UNIVERSITAT DE BARCELONA

Mecanismo de acción del factor represor Capicua, un sensor de señales Ras/MAPK

Marta Forés Maresma

TESIS DOCTORAL

Mecanismo de acción del factor represor Capicua, un sensor de señales Ras/MAPK

Marta Forés Maresma

Barcelona, Abril 2017

Departament de Genètica, Microbiologia i Estadística
Programa de Doctorat de Genètica
Facultat de Biologia
Universitat de Barcelona

Mecanismo de acción del factor represor Capicua, un sensor de señales Ras/MAPK

Memoria presentada por
Marta Forés Maresma
Para optar al grado de
Doctora por la Universitat de Barcelona

Esta Tesis Doctoral ha sido realizada en el Departamento de Biología del Desarrollo del Instituto de Biología Molecular de Barcelona (IBMB), perteneciente al Consejo Superior de Investigaciones Científicas (CSIC), Parc Científic de Barcelona (PCB), bajo la supervisión del Dr. Gerardo Jiménez Cañero.

El director	La alumna	El tutor
Dr. Gerardo Jiménez Cañero	Marta Forés Maresma	Dr. Ricard Albalat Rodriguez

ÍNDICE

ABREVIATURAS

Abreviatura	**Nombre completo**
ADN	Ácido Desoxirribonucleico
aos	*argos*
AP	Anteroposterior
ARN	Ácido Ribonucleico
ARNm	Ácido Ribonucleico mensajero
ATXN1	Ataxin-1
ATXN1L	Ataxin-1-like
bHLH	basic helix-loop-helix
Cas9	CRISPR Associate Nuclease 9
CBS	Capicua binding site
cic	*capicua*
cic-L	*capicua long*
cic-S	*capicua short*
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DA	Dorsoanterior
dpp	*decapentaplegic*
DUX4	Double Homeobox 4
DV	Dorsoventral
DYRK1A	Dual-specificity tyrosine-phosphorilation regulated protein kinase
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
eh1	Engrailed homology-1
EMSA	Electroforetic Mobility Shift Assay
en	*engrailed*
ERK	Extracellular signal-Regulated Kinase
ETS	E-Twenty Six
ETV	ETS variant
eve	*even-skipped*
FOXO4	Forkhead Box4
grk	*gurken*
gro	*groucho*
hkb	*huckebein*
HMG	High Mobility Group
ind	*intermediate neuroblasts defective*
kni	*knirps*
Kr	*Krüppel*
ISC	Intestinal Stem Cell
LEF	Lymphoid Enhancer binding Factor
MAPK	Mitogen-Activated Protein Kinase
mirr	*mirror*
mnb	*minibrain*
otd	*orthodenticle*
pb	pares de bases
PEA3	Polyoma Enhancer Activator 3

pnt	*pointed*
poliQ	poliglutamina
RTK	Receptor Tyrosine Kinase
Sox	SRY-related HMG-box
SRY	Sex-determining Region Y
TCF	T Cell-Specific Transcription Factor
TLE	Transducin Like Enhancer of Split
tll	*tailless*
tor	*torso*
trk	*trunk*
twi	*twist*
UAS	Upstream Activator Sequence
wap	*wings-apart*
wt	wild type
zen	*zerknüllt*

INTRODUCCIÓN

Las células de los organismos pluricelulares se comunican enviando y recibiendo señales entre ellas para coordinar su comportamiento y producir respuestas apropiadas.

Los mecanismos de comunicación se inician con moléculas secretadas extracelularmente, producidas por las células para comunicarse con sus vecinas o con células lejanas. La respuesta de las células receptoras a dichas señales depende de un elaborado sistema de proteínas que les permite responder a los mensajes procedentes de otras células. El sistema incluye receptores de la superficie celular, que unen la molécula señal, además de una gran variedad de proteínas señalizadoras que la transmiten intracelularmente. Al final de cada cascada de señalización intracelular se encuentran las proteínas diana que se modifican al activarse la vía, lo cual provoca cambios en el comportamiento celular. En los organismos pluricelulares, todos los aspectos del comportamiento celular como metabolismo, movimiento, proliferación y diferenciación son regulados por la señalización celular.

A menudo, el último componente de las vías de señalización son factores de transcripción que se encargan de controlar la expresión génica, de manera que la señal regula en última instancia el patrón espacio-temporal de la expresión de genes. De este modo, la señalización celular y la expresión regulada de factores de transcripción específicos son mecanismos fundamentales en el control de diversos procesos como la formación, identidad y patrón de tejidos y órganos durante el desarrollo de un animal. Los animales cuentan con un número relativamente pequeño de vías de señalización como las vías TGF-β/BMP, Notch, Wnt/β-catenin, Hedgehog y receptor tirosina quinasa (RTK) entre otras, pero éstas se utilizan reiteradamente durante el desarrollo.

Durante esta tesis se ha estudiado el mecanismo de acción del factor de transcripción Capicua (Cic), un represor transcripcional regulado por vías de señalización RTK con importantes funciones en el desarrollo de *Drosophila* e implicado en diferentes enfermedades en humanos.

1. Vías de señalización RTK

Las vías de señalización RTK regulan multitud de procesos celulares en todos los metazoos, incluyendo algunos tan importantes como proliferación, diferenciación,

supervivencia, metabolismo, migración y control del ciclo celular (Ullrich and Schlessinger, 1990). Además, una señalización anormal de la vía debido a cambios genéticos o una alteración de su actividad puede provocar múltiples enfermedades. De hecho, se han asociado mutaciones en los receptores RTK o una activación aberrante de las vías de señalización a enfermedades como cáncer, diabetes, inflamación y angiogénesis entre otras. Esta es la razón por la que se han desarrollado moléculas de nueva generación que bloquean o atenúan la actividad RTK para tratar algunos de estos trastornos.

Todos los RTKs tienen una estructura molecular parecida: un dominio de unión al ligando extracelular, una hélice transmembrana y una región citoplasmática con actividad tirosina quinasa. La activación de la vía se inicia por la unión del ligando al dominio extracelular del receptor que, oligomeriza y auto o trans-fosforila residuos de tirosina en su dominio intracelular. Los residuos de tirosina fosforilados son reconocidos por proteínas adaptadoras que inducen una cascada de fosforilación intracelular a través de diferentes vías como fosfolipasa C-γ, fosfatidilinositol 3-kinasa (PI3K) o la GTPasa pequeña Ras. Cuando Ras es fosforilada, continúa la vía llamada Ras/MAPK la cual sigue la señalización con la activación en serie de las proteínas quinasa Raf, MEK y MAPK (también llamada ERK) (Ullrich and Schlessinger, 1990). A menudo MAPK fosforila factores de transcripción nucleares provocando cambios en la actividad transcripcional (Figura 1). Los cambios en la actividad de estos factores se pueden dar de diferentes maneras, como por ejemplo cambiando su localización celular, expresión o estabilidad, o bien modulando su capacidad de remodelar la estructura de la cromatina o de unirse tanto a ADN como a co-reguladores (Whitmarsh, 2007).

En mamíferos, la vía Ras/MAPK fosforila un gran número de proteínas, principalmente factores de transcripción como Elk1, cFos o cMyc, los cuales regulan multitud de procesos. En *Drosophila*, los efectores de la vía mejor estudiados son los pertenecientes a la familia E-twenty six (ETS): el activador Pointed (Pnt) y el represor Yan (Brunner et al., 1994; Rebay and Rubin, 1995; Tootle and Rebay, 2005). No obstante, durante los últimos años se ha demostrado la importancia del factor HMG-box Cic como efector de dos vías RTK en *Drosophila*: la vía de Torso y la vía Epidermal

Growth Factor (EGF) en diferentes contextos (Jiménez et al., 2000; Goff et al., 2001; Roch et al., 2002; Astigarraga et al., 2007; Tseng et al., 2007; Andreu et al., 2012a). Además, también se ha demostrado la importancia de CIC, el homólogo de Cic en mamíferos, como efector de la vía EGFR que controla diferentes procesos implicados en la aparición de diferentes enfermedades (Dissanayake et al., 2011; Fryer et al., 2011; Lee et al., 2011; Okimoto et al., 2016).

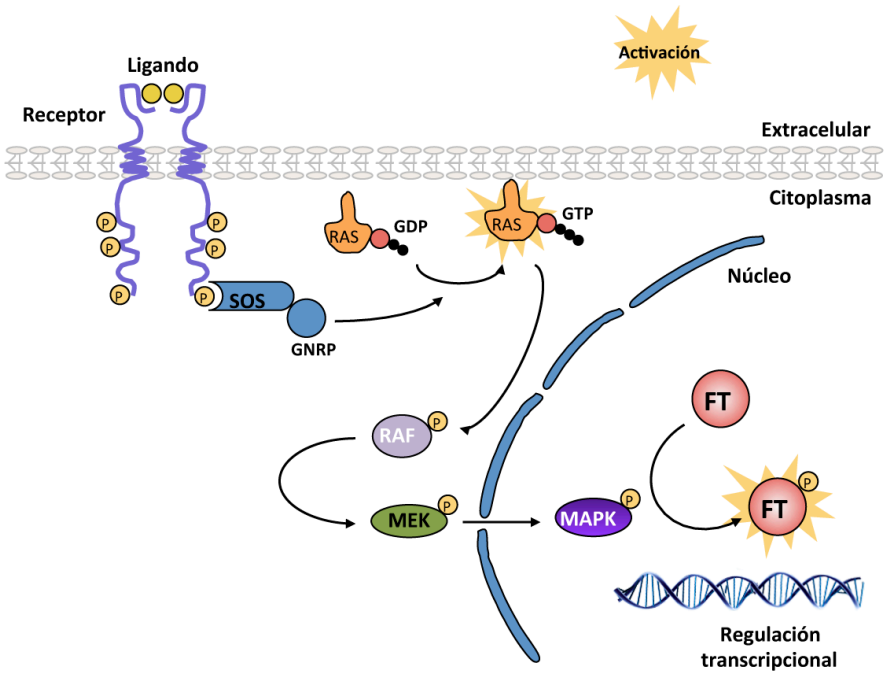


Figura 1. Esquema general de la vía RTK/Ras/MAPK. La vía se inicia con la unión del ligando al receptor el cual inicia la cascada de fosforilación que finalmente regula la función de factores de transcripción nucleares. FT: Factor de transcripción (Adaptado de Gilbert, 2003).

Desde el descubrimiento las formas oncogénicas de Ras y del receptor EGF (EGFR) hace más de 30 años, las vías RTK, y en concreto la vía Ras/MAPK, han sido ampliamente estudiadas hasta conocerse hoy en día muchos detalles de cómo la señal se transmite intracelularmente. Muchos de los componentes de esta vía están bien conservados a lo largo de los metazoos, así que no es sorprendente que la mayoría de factores inicialmente encontrados en especies como *Drosophila* o *C. elegans* tengan ortólogos conservados en humanos, mutaciones de los cuales están vinculadas a enfermedades como cáncer. Sin embargo, no se conoce del todo bien de qué manera la vía regula la expresión de genes en el núcleo de la célula. *Drosophila*, y en concreto la proteína Cic, ofrece un buen modelo para estudiar este aspecto y conocer cómo la

señalización por la vía regula la acción de factores de transcripción nucleares para controlar la expresión génica.

2. El factor represor Cic

2.1. Conservación evolutiva de Cic

Cic es una proteína HMG-box inicialmente identificada en *Drosophila* por su función por debajo de la vía RTK de Torso, que controla funciones esenciales en el desarrollo temprano del embrión (Jiménez et al., 2000). Esta proteína está conservada desde cnidarios hasta vertebrados y contiene 2 dominios claramente conservados: la HMG-box y el motivo C1. La HMG-box es el motivo de unión a ADN de Cic y es responsable también de la localización nuclear de la proteína. El motivo C1 está localizado en la región C-terminal y, aunque se desconoce su función molecular, se sabe que es imprescindible para la actividad represora de la proteína (Jiménez et al., 2000; Lee et al., 2002; Kawamura-Saito et al., 2006; Astigarraga et al., 2007). Existe un tercer dominio con una conservación menor con respecto a los dos anteriores llamado C2, a través del cual, en *Drosophila*, la MAPK Rolled se une a Cic para fosforilarla e inactivarla (Figura 2) (Astigarraga et al., 2007).

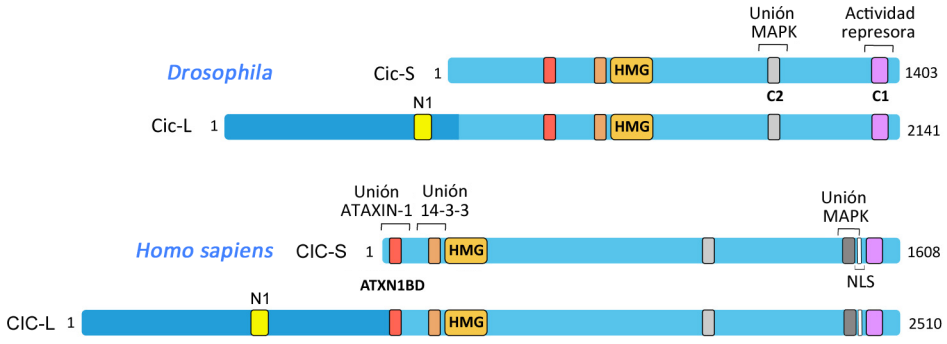


Figura 2. Características estructurales de las proteínas Cic de *Drosophila* y humanos. Las dos isoformas principales corta (Cic-S) y larga (Cic-L) están presentes en ambas especies. En *Drosophila* Cic-S ejerce casi todas las funciones conocidas de Cic. No se han encontrado diferencias funcionales entre las isoformas corta y larga en mamíferos. Los dominios funcionales de cada especie están indicados. Los números indican los aminoácidos de la proteína. NLS: Señal de localización nuclear. (Adaptado de Jiménez et al., 2012)

Tanto en *Drosophila* como en mamíferos existen al menos dos isoformas de Cic, una isoforma larga (Cic-L) y una corta (Cic-S), las cuales difieren en su región N-terminal siendo Cic-L la más larga. Hasta ahora, las funciones de la isoforma corta son las que se

han estudiado con más detalle. En *Drosophila,* Cic-L tiene un papel específico en oogénesis y se sospecha que es necesario para la localización y migración celular de productos génicos durante este proceso (Rittenhouse and Berg, 1995; Jiménez et al., 2000; Goff et al., 2001; Roch et al., 2002; Dorman et al., 2004; Astigarraga et al., 2007), y aunque su función concreta no se conoce, se sabe que en la región N-terminal existe otro motivo conservado, llamado N1, del que también se desconocen aún sus funciones moleculares (Figura 2) (Lam et al., 2006).

Los factores HMG-box como Cic son proteínas nucleares con funciones diversas en la célula. La HMG-box, su motivo de unión a ADN, tiene unos 75 aminoácidos y consiste en 3 hélices de tipo α con un plegamiento característico en forma de L. Estos factores se unen al surco estrecho de la doble hélice de ADN a través de interacciones electrostáticas e hidrofóbicas, provocando un estrechamiento del surco estrecho y el doblamiento de la doble hélice hacia el surco mayor (Reményi et al., 2003). Los factores HMG-box se pueden clasificar en dos grandes grupos: los que se unen a ADN de manera inespecífica de secuencia y los que lo hacen de manera específica. Las proteínas HMG-box con inespecificidad de secuencia, suelen tener más de un dominio HMG-box y funcionan como subunidades de complejos de remodelación de cromatina. Por otro lado, las proteínas con un solo motivo HMG-box como Sox/SRY (Sox) o TCF/LEF-1 (TCF), se unen a lugares específicos de promotores o *enhancers*, y funcionan como factores de transcripción reguladores del desarrollo (revisado en Štros et al., 2007; Malarkey and Churchill, 2012; Kamachi and Kondoh, 2013). A pesar de que la HMG-box de Cic es muy similar en secuencia a la HMG-box de la subfamilia Sox, el nivel de similitud es insuficiente para considerarlo un miembro de esta subfamilia, y se ha definido como miembro de la subfamilia de HMG-box relacionadas con Sox (Lee et al., 2002).

Cic ejerce su represión transcripcional uniéndose a secuencias octaméricas específicas T(G/C)AATG(A/G)A en las regiones reguladoras de sus genes diana a través de la HMG-box tanto en mamíferos como en *Drosophila* (Kawamura-Saito et al., 2006; Lam et al., 2006; Löhr et al., 2009; Kazemian et al., 2010; Ajuria et al., 2011; Lee et al., 2011). En *Drosophila*, la represión mediada por Cic está relacionada con el control de la transcripción dependiente de RTK, ya que Cic reprime genes que están inducidos por la

vía, y esta inducción ocurre, en parte, a través de la disminución de la represión por Cic (Jiménez et al., 2000; Goff et al., 2001; Roch et al., 2002; Atkey et al., 2006; Astigarraga et al., 2007; Tseng et al., 2007; Ajuria et al., 2011). En mamíferos la represión mediada por CIC está regulada de una manera similar por vías RTK (Dissanayake et al., 2011; Fryer et al., 2011; Lee et al., 2011; Kim et al., 2015).

2.2. Mecanismos de represión por Cic, correpresores: Groucho y Ataxin-1

Los mecanismos por los que Cic se une a ADN y reprime sus genes diana no se conocen del todo bien, pero hay dos correpresores que están implicados en este proceso.

Por un lado, se ha sugerido que, en *Drosophila*, la actividad de Cic depende del correpresor Groucho (Gro). Gro pertenece a la familia de correpresores en la que se incluyen las proteínas de vertebrados Transducin-Like Enhancer of Split (TLE) y Groucho related gene, que operan en muchas vías de señalización. Gro está implicado en muchos contextos del desarrollo de *Drosophila*, como en la segmentación del embrión, neurogénesis, determinación del sexo, determinación del eje dorsoventral (DV) y la formación del patrón del ala (Paroush et al., 1994; Paroush et al., 1997; Parkhurst, 1998; Chen and Courey, 2000; Roch et al., 2002), y los homólogos en vertebrados actúan durante procesos como neurogénesis, osteogénesis y hematopoyesis entre otros (revisado en Gasperowicz and Otto, 2005).

Gro se considera un correpresor general bien establecido: no interacciona con ADN directamente, sino que es reclutado a las regiones reguladoras de los genes diana por factores de transcripción de unión a ADN. En cambio, no está claro el mecanismo por el que Gro actúa inhibiendo la transcripción. Una de las maneras en la que los correpresores pueden llevar a cabo su función es modificando el estado de la cromatina (Struhl, 1998) y se ha sugerido que Gro podría estar utilizando un mecanismo similar (Chen et al., 1999; Chen and Courey, 2000).

Otro de los aspectos claves del mecanismo de acción de Gro es la manera como interacciona con sus ligandos. Por cristalografía de rayos X se ha visto que los represores se unen a Gro a través de su dominio C-terminal WD el cual forma una estructura de 7 aspas alrededor de un poro central llamada *β-propeller* (TLE1 humano; Pickles et al., 2002), una estructura conocida por mediar muchas interacciones

proteína-proteína (revisado en Li and Roberts, 2001). Muchos estudios indican que el dominio WD de Gro interacciona con motivos peptídicos cortos de diferentes familias de factores de transcripción. Estos motivos incluyen el motivo C-terminal WRPW de los factores de transcripción bHLH de la familia de Hairy, que incluye Hairy, Deadpan y Enhancer of Split (Paroush et al., 1994; Fisher et al., 1996; Jiménez et al., 1997) y el motivo interno Engrailed homology-1 domain (eh1) cuya secuencia peptídica es FxIxxIL y se encuentra, por ejemplo, en Engrailed, Dorsal, Goosecoid y Oddskipped (Smith and Jaynes, 1996; Dubnicoff et al., 1997; Jiménez et al., 1997; Tolkunova et al., 1998; Jiménez et al., 1999; Copley, 2005). Además de los motivos represores WRPW y eh1, Gro también reconoce péptidos parecidos a WRPW, como el motivo interno FRPW de Huckebein (Hkb) (Goldstein et al., 1999; Jennings et al., 2006) y el motivo WRPY de Runx (Canon and Banerjee, 2003; Jennings et al., 2006). Análisis funcionales y estructurales han definido que los motivos WRPW y eh1 adoptan diferentes conformaciones espaciales, pero solapan en algunos sitios en la ocupación del poro central de la estructura *β-propeller* del motivo WD de Gro (Jennings et al., 2006).

A pesar de que Cic no tiene ningún motivo conocido de interacción con Gro, se ha visto que, durante la embriogénesis de *Drosophila*, pérdidas de función de *cic* y *gro* dan efectos similares, siendo necesaria la presencia del correpresor para la represión de los genes diana de Cic *tailless* (*tll*) y *huckebein* (*hkb*) (explicado con más detalle en el apartado 3.2.1 de la introducción) (Paroush et al., 1997; Jiménez et al., 2000; Cinnamon et al., 2008; Jennings et al., 2008). Además, la unión de Cic al *enhance*r de *hkb* correlaciona con la asociación de Gro a este *enhancer* (Ajuria et al., 2011). Aunque estos datos sugieren un modelo donde Gro actúa como correpresor de Cic, los enlaces moleculares entre estos dos factores no se conocen y no existen evidencias de interacciones físicas entre ambas proteínas *in vivo* (Ajuria et al., 2011).

No obstante, hasta ahora, no existe evidencia de que el mecanismo represor de Cic a través de Gro esté conservado en otras especies, incluidos los humanos.

Por otro lado, varios estudios indican que la represión mediada por CIC en mamíferos implica la formación de complejos con el correpresor Ataxin-1 (ATXN1) y el factor relacionado Ataxin-1-like (ATXN1L) (Lam et al., 2006; Bowman et al., 2007; Lim et al., 2008; Crespo-Barreto et al., 2010; Lee et al., 2011). ATXN1 interacciona con el motivo

ATXN1-BD de CIC para formar complejos, el cual está moderadamente conservado evolutivamente (Figura 2) (Lam et al., 2006). La presencia de ATXN1 y ATXN1L estabiliza la proteína CIC, ya que los niveles de ésta se ven reducidos en fondos mutantes para ambos genes, mientras que los niveles de ARNm se mantienen (Lam et al., 2006; Lee et al., 2011). Además, la co-expresión de ATXN1 o ATXN1L con CIC aumenta la actividad represora de CIC *in vitro*, sugiriendo que los complejos ATXN1/ATXN1L-CIC podrían actuar cooperativamente para regular la expresión de los genes diana de CIC (Lee et al., 2011). De todos modos, no hay ninguna evidencia que sugiera que en *Drosophila,* Cic medie represión a través de Atx-1, el homólogo de ATXN1 en esta especie.

3. Funciones de Cic en *Drosophila*

En esta tesis, se ha utilizado *Drosophila melanogaster* para evaluar el mecanismo de acción de Cic, examinando con detalle su función en diferentes tejidos y analizando la capacidad de reprimir sus genes diana en diferentes condiciones y diferentes momentos del desarrollo.

Drosophila es un organismo holometábolo que desarrolla una metamorfosis completa. Después de la deposición del huevo por la hembra, el cigoto comienza a dividirse y empieza la embriogénesis. En este momento el embrión se convierte en larva (primer estadio larvario o L1) que en 24 horas eclosiona del huevo y empieza a sumergirse en la comida para alimentarse y crecer, pasando por 2 estadios larvarios más (L2 y L3) hasta que pupa y realiza la metamorfosis convirtiéndose en mosca adulta al cabo de 4 días. Durante la metamorfosis, muchos de los tejidos larvarios son destruidos. Los tejidos adultos como las alas, las patas y los ojos se desarrollan a partir de grupos de células llamados discos imaginales, que son primordios de las estructuras adultas.

Cic actúa en diferentes momentos del desarrollo de la mosca por debajo de las vías RTK de Torso y EGFR. Por debajo de la vía de Torso, Cic está inactivado y actúa en el embrión para establecer el sistema DV y el sistema terminal de este tejido. Por otro lado, la vía EGF controla múltiples eventos de patrón y especificación celular en *Drosophila* (Shilo, 2003), muchos de los cuales están mediados por Cic. En *Drosophila*,

la regulación negativa de Cic por parte de la vía EGF se da en múltiples tejidos: en el ovario para el establecimiento del eje DV del futuro embrión y para el patrón de los apéndices respiratorios del huevo, en el neuroectodermo del embrión para regular la diferenciación de neuroblastos y en el disco imaginal del ala para un correcto patrón de la venación del ala.

3.1. Funciones de Cic en las células foliculares

El oocito maduro de *Drosophila* es una célula altamente compleja que contiene la información para establecer los ejes anteroposterior (AP) y DV del futuro embrión.

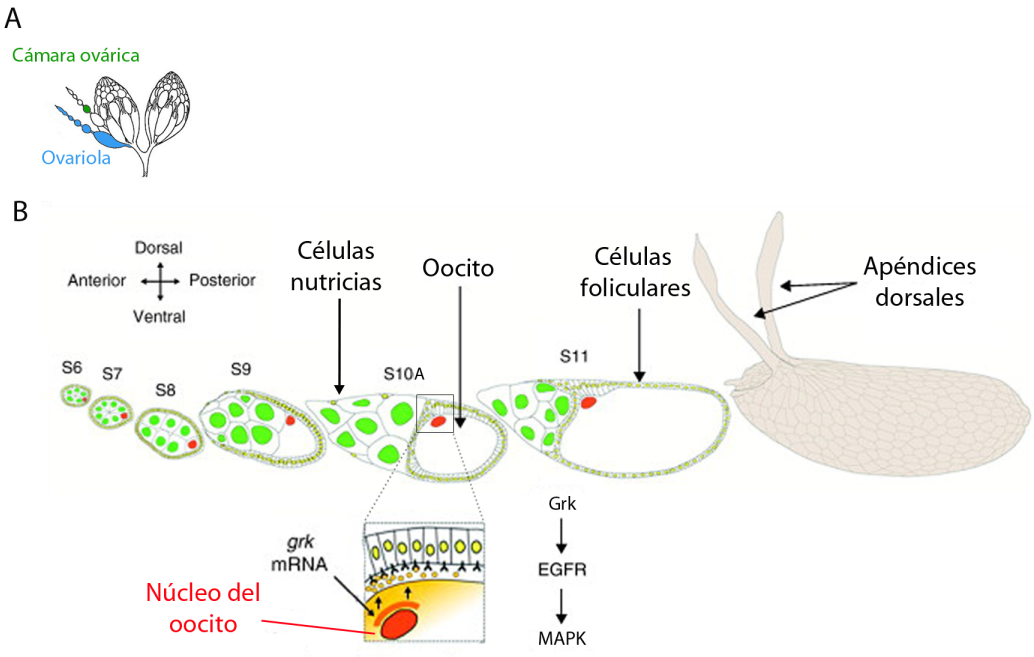


Figura 3. Del ovario al huevo de *Drosophila*. (A) Representación de los ovarios de *Drosophila* y su organización. **(B)** Representación esquemática de los estadios 6-11 de la oogénesis hasta llegar al huevo. En el estadio 10 el ARNm de *grk* se acumula cerca del núcleo del oocito. La proteína se produce localmente en el oocito y se secreta en el espacio perivitelino, donde difunde para unirse a receptores de las células foliculares que lo envuelven. La activación de EGFR por Grk inicia la cascada Ras/MAPK. (Adaptado de Cheung et al., 2011)

El ovario de *Drosophila* consiste en 16-20 estructuras tubulares denominadas ovariolas, las cuales contienen cámaras ováricas en diferentes estadios de maduración (Figura 3A) (King, 1970). Estas cámaras ováricas contienen el oocito, el cual se desarrolla junto a quince células hermanas denominadas células nutricias y está envuelto por una monocapa de células foliculares (Figura 3B) (Spradling, 1993). Tanto las células nutricias como las foliculares proporcionan al oocito constituyentes citoplasmáticos, además de determinantes específicos necesarios para la formación

del patrón inicial del embrión. Los determinantes de las células nutricias (en su mayoría ARNm) se acumulan de forma asimétrica en el oocito y se activan en el momento de la fertilización. Por otro lado, las células foliculares secretan factores de desarrollo que se asocian a la superficie extracelular del embrión. Estos factores, mantienen la información posicional generada en la oogénesis para la especificación del patrón DV y de los extremos del embrión a través de la generación de señales que activan receptores transmembrana en posiciones restringidas (Figura 3B) (St Johnston and Nüsslein-Volhard, 1992).

3.1.1. Establecimiento del eje DV en el oocito

Durante la oogénesis, el oocito crece y envía señales a las células foliculares que adquieren una polaridad a lo largo de los ejes AP y DV. Durante los estadios 7-8 de la oogénesis, el núcleo del oocito migra a una posición cortico-anterior especificando en esa zona la futura región dorso-anterior (DA) del huevo. El posicionamiento asimétrico del núcleo del oocito dirige el transporte del tránscrito de *gurken* (*grk*) a la esquina DA del oocito. Durante el estadio 9-10, el producto Grk es secretado desde el oocito y activa el receptor EGFR en las células foliculares DA (Figura 3B) (revisado en Cheung et al., 2011). La activación de EGFR señaliza a través de la vía Ras-MAPK induciendo la expresión de *mirror* (*mirr*), el cual codifica para un factor de transcripción con homeodominio la regulación del cual tiene un papel fundamental en el establecimiento de la polaridad DV en las células foliculares (Jordan et al., 2000; Zhao et al., 2000; Andreu et al., 2012a) y en la especificación de las células DA que dirigen la formación de los apéndices respiratorios en el huevo (Figura 3B) (Berg, 2005; Atkey et al., 2006). Mutaciones que previenen la señalización EGFR llevan a la ventralización del huevo y del embrión, mientras que la activación ectópica de la vía los dorsaliza (Schüpbach, 1987; Queenan et al., 1997).

En las células foliculares, Cic reprime la expresión de *mirr* (Atkey et al., 2006; Andreu et al., 2012b). Por lo tanto, en las células DA la vía EGFR por un lado induce la expresión de *mirr*, y por otro lado fosforila e inactiva Cic (Figura 4A) (Goff et al., 2001), provocando una relocalización parcial de la proteína al citoplasma, reduciendo sus niveles un 50% (Figura 4B) (Astigarraga et al., 2007). La inhibición de Cic por la vía contribuye a la inducción de la transcripción de *mirr* a través de desrepresión. La

inducción de *mirr* solo en el dominio más anterior, es debido al papel de Decapentaplegic (Dpp), el cual ayuda a su inducción en esta región de las células foliculares (Goff et al., 2001; Atkey et al., 2006). El producto Mirr tiene la función de mantener el dominio ventral de *pipe* reprimido en las células DA (Figura 4A) (Andreu et al., 2012a; Andreu et al., 2012b; Fuchs et al., 2012; Technau et al., 2012). Pipe es una sulfotransferasa, cuya activación en la región ventral del epitelio folicular determinará la región ventral del embrión temprano (Sen et al., 1998). De este modo, la inactivación de Cic por la vía EGFR modula la distribución espacial de *mirr* en las células foliculares laterales, lo que contribuye a definir la posición en la que el borde de expresión de *pipe* se forma (Andreu et al., 2012b). Consiguientemente, perdidas de la función de *cic* causan una desrepresión de *mirr* hacia las células foliculares ventrales (Figura 4C,D) y por lo tanto causan la perdida de expresión de *pipe* (Figura 4E,F). Estos efectos llevan a una dorsalización del embrión, por lo que los embriones producidos por hembras mutantes *cic* no tienen estructuras ventrales como los dentículos ventrales y están compuestos exclusivamente por epidermis dorsal (Figura 4G,H) (Goff et al., 2001; Atkey et al., 2006).

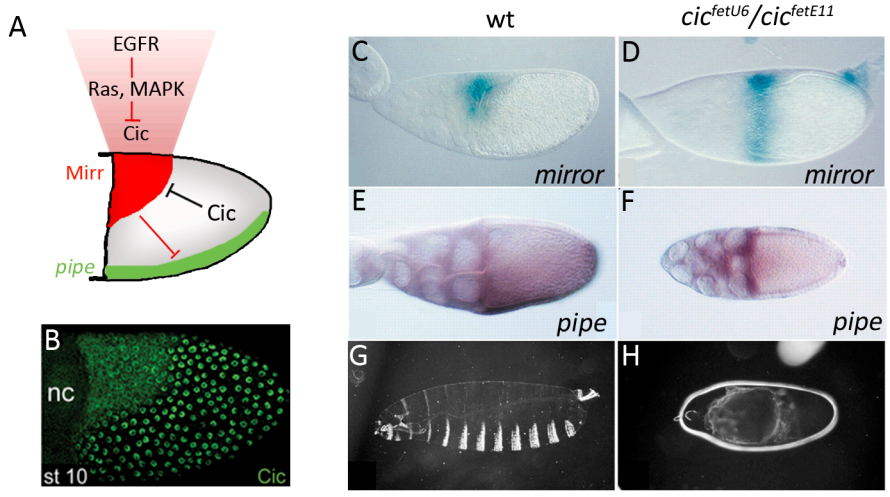


Figura 4. La acción de Cic es necesaria para el establecimiento del eje DV en el ovario. (A) Esquema general de la función de Cic por debajo de la vía de EGF en las células foliculares del ovario. La vía induce la expresión de *mirr* a la vez que regula negativamente a Cic el cual colabora en mantener la expresión de *mirr* en las células DA y mantener la expresión ventral de *pipe*. **(B)** Inmunotinción contra la proteína Cic de una cámara ovárica; en la región DA hay una relocalización de Cic al citoplasma debido a la fosforilación por la vía **(C,D)** Detección del ARNm de *mirr* en cámaras ováricas *wild type* (*wt*) (C) y mutante para *cic* en la que expresión de *mirr* se desreprime hacia ventral (D). **(E,F)** Detección del ARNm de *pipe* en cámaras *wt* (E) y mutante para *cic* donde el dominio ventral de *pipe* desaparece a causa de la desrepresión de *mirr* (F). **(G,H)** Cutículas de embriones *wt* (G) y mutante *cic* donde el embrión está dorsalizado (H). Las cámaras ováricas y los huevos están orientados con anterior a la izquierda, posterior a la derecha, dorsal arriba y ventral abajo. (Adaptado de Goff et al., 2001)

3.1.2. Regulación del patrón de los apéndices dorsales

Al final de la oogénesis, las células foliculares secretan la cubierta del huevo, el corion, el cual muestra asimetrías axiales pronunciadas en su forma y en la presencia de estructuras especializadas como los apéndices respiratorios, que flanquean la línea media dorsal (Figura 3B) (Waring, 2000).

La formación de los apéndices está controlada por el gradiente de señal EGFR iniciado por Grk, donde los niveles más altos de señal corresponden a la región más dorso-anterior del oocito y dan lugar a la activación de Pointed (Pnt) que establece la línea media dorsal (el espacio que separa la formación de los dos apéndices) (Boisclair Lachance et al., 2009). Por otro lado, niveles más bajos de señalización, dan lugar a la formación de los apéndices a través de la inducción de la expresión de *mirr* (Figura 5A) (Berg, 2005; Atkey et al., 2006; Boisclair Lachance et al., 2009). De este modo, niveles mayores de señal EGFR producen un dominio más ancho de expresión de *pnt* y por lo tanto apéndices dorsales más separados (Neuman-Silberberg and Schupbach, 1994; Boisclair Lachance et al., 2009), mientras que niveles reducidos de esta señal provocan la pérdida del destino celular dorsal causando la fusión de los apéndices (Schüpbach, 1987).

A pesar de que no se conoce del todo cómo la señal EGFR controla la expresión de los genes diana en el núcleo de las células foliculares en el proceso de especificación de los apéndices dorsales, un mecanismo directo implica la inhibición de Cic a través de la fosforilación por MAPK (Astigarraga et al., 2007). Tal como ocurre en la especificación DV en el ovario, Cic se requiere de forma autónoma celular en las células foliculares ventrales y laterales, en este caso para reprimir los destinos celulares de formación de apéndices, y la transformación a este tipo de células formadoras de apéndices coincide y depende de la expresión de *mirr*. De acuerdo con esto, la expresión ectópica de la proteína Mirr es suficiente para producir material formador de apéndices en cualquier posición del epitelio folicular (Atkey et al., 2006). En cambio, pérdidas de función de *cic* provocan una desrepresión de *mirr* en la región anterior del epitelio folicular (Figura 4D), causando unos apéndices dorsales más gruesos que están localizados más lateralmente, con un espacio dorsal entre los dos apéndices mayor y con material tipo apéndice extra rodeando la circunferencia anterior del huevo (Figura 5B,C) (Goff et al.,

2001; Atkey et al., 2006). Por el contrario, formas de Cic insensibles a la vía producen una proteína Cic constitutivamente activa, lo que resulta en la fusión de los apéndices respiratorios tal como ocurre cuando se reduce la señal EGFR (Astigarraga et al., 2007).

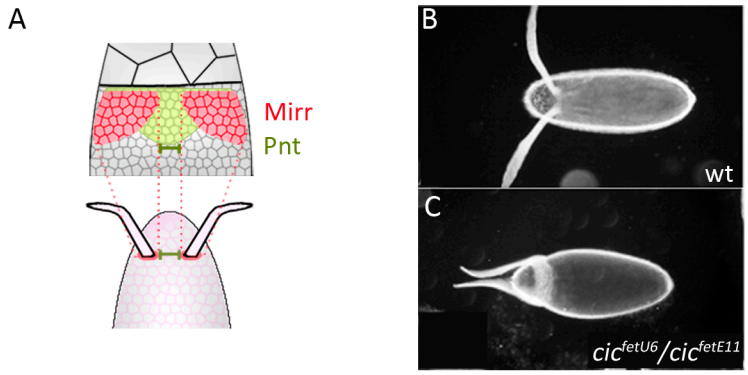


Figura 5. Cic está implicado en el patrón de los apéndices respiratorios del huevo. (A) Esquema general del mecanismo de formación de apéndices. La señalización EGFR induce *mirr* marcando el dominio donde se formarán los apéndices dorsales, y Pnt coincidiendo con el pico máximo de señal EGFR, el cual establecerá la línea media dorsal que marcará la separación entre apéndices. **(B,C)** Huevos depositados por hembras *wt* (B) o hembras mutantes *cic* los cuales tienen los apéndices dorsales más separados y material formador de apéndice rodeando toda la región anterior del huevo (C). (Adaptado de Goff et al., 2001)

3.2. Funciones de Cic en el embrión

La información posicional generada durante la oogénesis se mantiene durante el desarrollo del embrión temprano y lleva al establecimiento de los ejes del embrión. Las primeras fases del desarrollo del embrión dependen de actividades génicas codificadas en la hembra y que son contribuidas al embrión en forma de moléculas de ARNm y proteínas de origen materno (depositadas a lo largo de la oogénesis). Estos genes son los denominados genes de efecto materno, y mutaciones en ellos no afectan a la hembra portadora sino a su progenie. A lo largo del eje AP, el embrión se divide en diferentes regiones que darán lugar al tórax, el abdomen y las estructuras terminales de la larva (Figura 6A). Hay tres clases de genes maternos que especifican el eje AP: los que especifican las regiones anteriores; los que especifican las regiones posteriores; y los que especifican las regiones terminales del embrión (St Johnston and Nüsslein-Volhard, 1992). La actividad localizada de los genes maternos, los cuales generan gradientes de proteínas morfógenas, generan un patrón espacial específico de la expresión de genes cigóticos. Hay 4 clases de genes cigóticos actuando a lo largo del eje AP de manera jerárquica. Los genes *gap*, que están directamente regulados por los genes maternos y que especifican regiones anchas que se redefinirán por la acción de los genes *pair-rule*, resultando en un patrón periódico de expresión génica. Los genes *segment polarity*, elaboran el patrón entre segmentos. Las acciones de estas tres

clases de genes determinan el dominio espacial de genes homeóticos Hox que definen las identidades de cada segmento (Figura 6A).

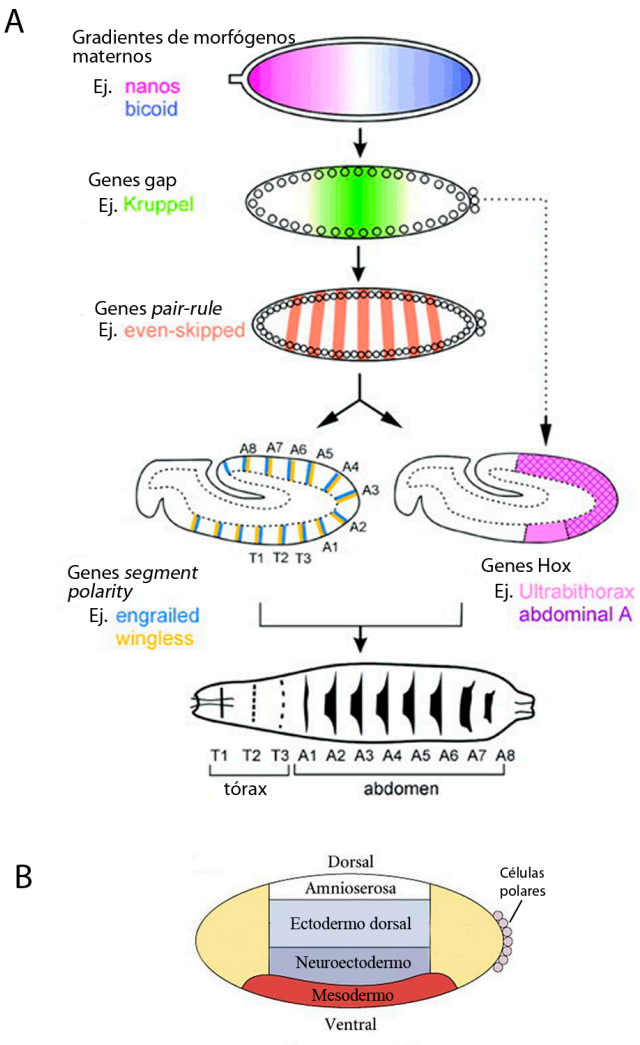


Figura 6. Establecimiento de los patrones AP y DV en el embrión de *Drosophila*. (A) El patrón AP se establece por el efecto de genes maternos que forman gradientes de proteínas morfógenas. Estos determinantes morfógenos regulan la activación de los genes *gap*, los cuales definen territorios amplios en el embrión. Los genes *gap* regulan la expresión de los genes *pair-rule*, cada uno de los cuales divide el embrión en regiones de unos dos segmentos. Los genes *segment polarity* dividen el embrión en segmentos a lo largo del eje AP. Las acciones de estos genes definen los dominios espaciales de los genes Hox que definen las identidades de cada segmento de la futura larva. **(B)** Esquema de la subdivisión del embrión en el eje DV. (Adaptado de Sanson, 2001)

Además de los genes de segmentación, existen una serie de genes cigóticos que no intervienen en procesos de formación de segmentos del embrión. Es el caso de los genes cigóticos que regionalizan el embrión en su eje DV. El eje DV se divide en 4 regiones en la embriogénesis temprana de ventral a dorsal: el mesodermo, neuroectodermo, el ectodermo dorsal y la amnioserosa. El mesodermo formará músculos y otros tejidos conectivos internos; el neuroectodermo dará lugar al sistema nervioso de la larva y el ectodermo dorsal a la epidermis de la larva. Finalmente, la amnioserosa es una membrana extraembrionaria en la cara dorsal del embrión (Figura 6B).

3.2.1. Regulación del sistema terminal

El grupo de los genes terminales está compuesto por genes que especificaran las estructuras en los extremos del embrión al final de la embriogénesis: el *labrum* y las estructuras mandibulares en anterior y los segmentos abdominales 7º y 8º, los espiráculos, el telson y las estructuras anales en posterior (Figura 7) (Schüpbach and Wieschaus, 1986; Jurgens and Hartenstein, 1993). La especificación de las dos regiones terminales depende de la expresión localizada de los genes *gap tll* y *hkb* en respuesta a la señal de Torso en los polos del embrión (Pignoni et al., 1990; Duffy and Perrimon, 1994; Brönner and Jäckle, 1996).

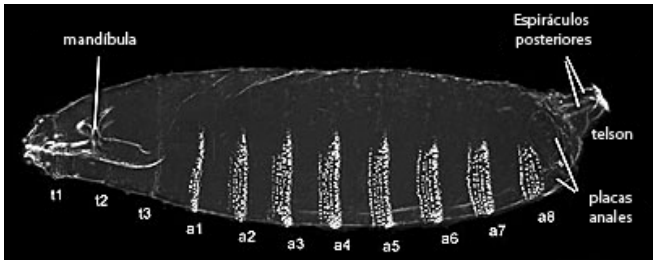


Figura 7. Cutícula de un embrión de *Drosophila*. La cutícula es producto de la secreción de las células epidérmicas. La zona ventral muestra dentículos en el borde anterior de cada segmento (3 torácicos y 8 abdominales). En la región anterior se muestran las estructuras mandibulares y en la posterior los espiráculos, el telson y las placas anales. (Adaptado de FlyMove)

Torso es un receptor RTK heredado por vía materna y distribuido a lo largo de toda la membrana plasmática del embrión temprano pero activado solamente en los polos por su ligando Trunk (Trk) (Schüpbach and Wieschaus, 1986; Casanova et al., 1995; Casali and Casanova, 2001). La expresión de Torso-like en los polos de la cámara ovárica y su acumulación en los polos del embrión, son necesarios para la activación de Torso mediada por Trk de manera localizada en los polos (Stevens et al., 1990; Martin et al., 1994; Stevens et al., 2003; Johnson et al., 2015; Mineo et al., 2015).

La activación de Torso en los polos del embrión induce una cascada de fosforilación a través de la vía Ras-MAPK que finalmente fosforila y degrada a Cic en el núcleo. Aún no se conoce el mecanismo por el que Cic es degradado, pero se sabe que Rolled (la MAPK de *Drosophila*) se une a Cic a través de su motivo C2, fosforilando y degradando la proteína y provocando un gradiente de Cic ascendente hacia el centro del embrión (Figura 8A-C).

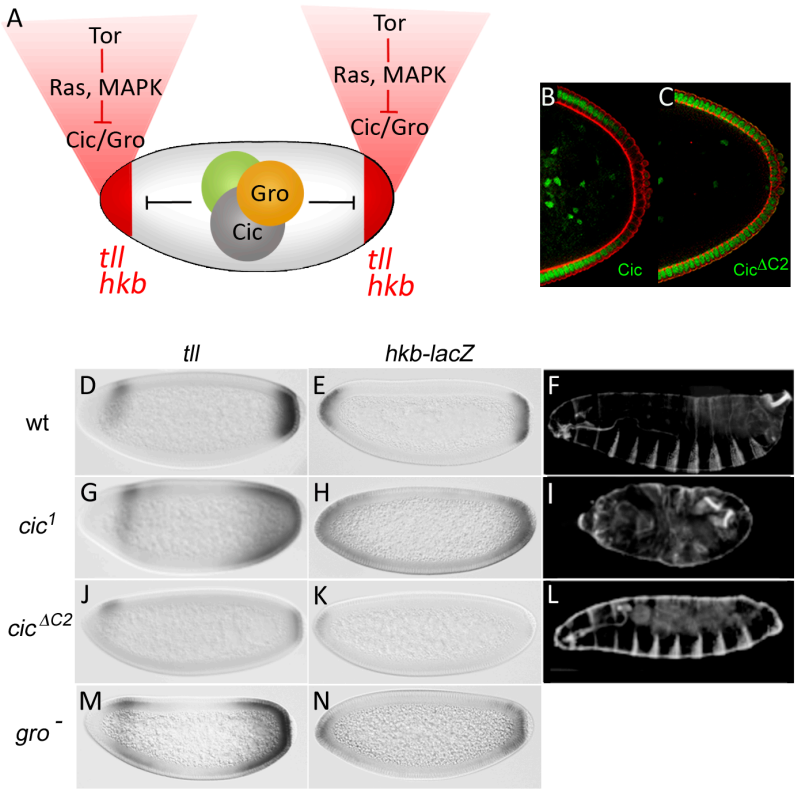


Figura 8. La vía de Torso regula la actividad de Cic en la represión asistida por Gro durante la especificación del sistema terminal. (A) Esquema general de la represión mediada por Cic regulado por la vía de Torso en el embrión temprano. La vía inhibe la represión mediada por Cic y Gro en los extremos del embrión permitiendo la expresión localizada de *tll* y *hkb* **(B,C)** Inmunotinciones contra Cic en un embrión *wt* donde se ve la degradación de Cic en el polo posterior en respuesta a la vía (B) y un embrión portador de un transgén de Cic que carece del motivo de anclaje de MAPK, C2; esta forma de Cic es insensible a la vía y no es degradado en el polo posterior (C). **(D-N)** Detección del ARNm de *tll* y *hkb,* y cutículas en diferentes embriones: *wt,* donde la expresión de los genes diana está restringida a los polos (D-F); cic^{1}, un alelo hipomorfo para la forma Cic-S, donde la expresión de los genes diana esta desreprimida (G-H) y en la cutícula se observa una expansión de las estructuras terminales a expensas de los segmentos abdominales (I); $cic^{\Delta C2}$, que carece del motivo de unión a MAPK y donde los genes diana están reprimidos incluso en los polos del embrión (J-K), causando una pérdida de las estructuras terminales (L); y alelos nulos de *gro:* gro^{E48} (M) y gro^{MB36} (N) dónde también se aprecia la desrepresión de los genes diana. De aquí en adelante, la orientación de los embriones será con anterior a la izquierda, posterior a la derecha, dorsal arriba y ventral abajo. (Adaptado de Astigarraga et al., 2007; Ajuria et al., 2011)

La degradación de Cic por la vía es imprescindible para una correcta formación de las regiones terminales del embrión (Figura 8A-C) (Astigarraga et al., 2007). La expresión de Cic en las regiones centrales del embrión reprime la expresión de sus genes diana *tll* y *hkb*, y mantiene su expresión localizada en los polos (Jiménez et al., 2000; Astigarraga et al., 2007; Ajuria et al., 2011). De este modo, la vía de Torso induce la expresión de *tll* y *hkb* en los polos del embrión a través de inhibición de la acción de Cic (Figura 8A,D,E). Tal como se ha comentado anteriormente, Gro también juega un papel importante en la represión de los genes diana de Cic, ya que tanto pérdidas de función

de *cic* o de *gro* resultan en la desrepresión *tll* y *hkb* (Figura 8A,G,H,M,N) (Paroush et al., 1997; Jiménez et al., 2000; Cinnamon et al., 2008; Jennings et al., 2008). Tll a su vez mantiene reprimida en la región posterior la banda central de expresión de los genes *gap knirps* (*kni*) y *Krüppel* (*Kr*) (Steingrimsson et al., 1991), de manera que pérdidas de función de *cic* llevan a la desaparición del dominio central de *kni* y *Kr*, lo que resulta en la pérdida de los segmentos abdominales, causando lo que se conoce como un fenotipo *capicua*, donde los embriones solo tiene definidas las regiones más anteriores y posteriores (Figura 8I) (Jiménez et al., 2000). Por el contrario, formas de Cic que carecen del motivo C2 producen una proteína Cic constitutivamente activa e incapaz de ser regulada por la señal, lo que resulta en fenotipos de ganancia de función en los que se pierde la expresión de los genes *gap tll* y *hkb,* y consecuentemente la pérdida de las estructuras terminales del embrión (Figura 8C, J-L) (Astigarraga et al., 2007).

3.2.2. Establecimiento del eje DV en el embrión

La señalización de Torso en los polos del embrión también regula procesos que operan durante el patrón DV (Casanova, 1991; Rusch and Levine, 1994). Este patrón depende de la expresión del morfógeno Dorsal en el embrión, la cual depende de eventos iniciados en las células foliculares. La expresión del factor Pipe en la región ventral de las células foliculares modifica componentes estructurales de la membrana vitelina necesarios para la activación del receptor Toll en las células ventrales del embrión, lo que provoca la translocación del factor Dorsal al núcleo formando un gradiente a lo largo del eje DV (Roth et al., 1989; Rushlow et al., 1989; Steward, 1989). De este modo, Dorsal se acumula en la región ventral del embrión y actúa como activador y represor de la transcripción: activa genes específicos ventrales como *twist* (*twi*) o *snail* y reprime genes específicos dorsales como *zerknüllt* (*zen*), *tolloid* y *dpp* (Ip et al., 1991; Jiang et al., 1991; Kirov et al., 1993; Rusch and Levine, 1996). La actividad represora de Dorsal sobre *zen*, requiere la presencia de varios factores como Gro, Dead-Ringer, Cut y Cic (Dubnicoff et al., 1997; Valentine et al., 1998; Jiménez et al., 2000), de manera que Cic participa en la función represora de Dorsal, pero no en la activadora, ya que mutantes *cic*, mantienen normal la expresión de *twi*, mientras que *zen* se encuentra desreprimido (Jiménez et al., 2000).

3.2.3. Regulación de la diferenciación de neuroblastos

La activación localizada de EGFR en el neuroectodermo del embrión regula la diferenciación de neuroblastos induciendo la expresión del gen *intermediate neuroblasts defective* (*ind*) que codifica para un factor de transcripción homeobox requerido para el patrón del cordón nervioso (Weiss et al., 1998; Von Ohlen and Doe, 2000). La expresión de *ind* requiere el alivio de su represión por Cic, a través de la inhibición por la vía de EGF (Figuras 6,9) (Ajuria et al., 2011).

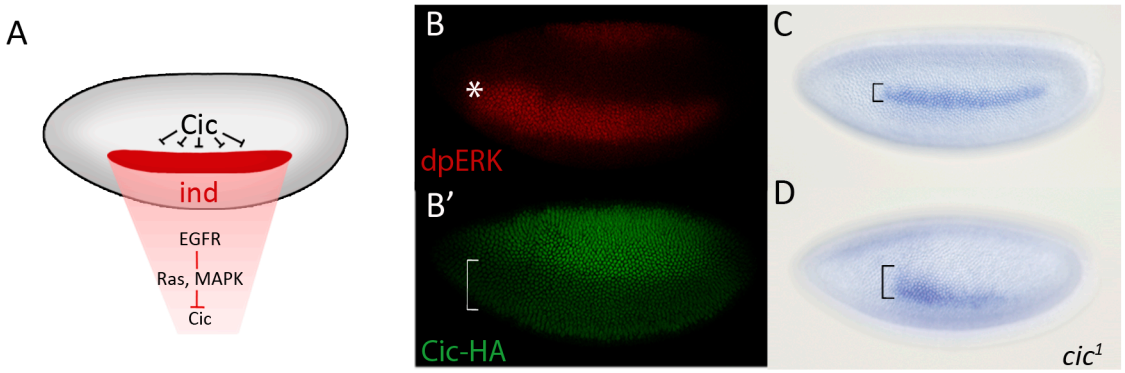


Figura 9. Cic participa en la regulación de la diferenciación de neuroblastos en el embrión. (A) Esquema general de la expresión de *ind* inducida por la vía EGFR a través de la inhibición de Cic. **(B,B')** Inmunotinción contra MAPK fosforilada (dpERK) (B) y contra HA en un embrión portador de un transgén de Cic marcado con HA (B'). Se observa que la banda donde la vía está activa solapa con la disminución de los niveles de Cic-HA. **(C,D)** Detección del ARNm de *ind* de embriones *wt* (C) y mutante cic^{1} donde se aprecia la desrepresión de *ind* (D). (Adaptado de Ajuria et al., 2011)

3.3. Regulación del patrón de venación del ala

El ala de *Drosophila* consiste en dos capas epidérmicas, una dorsal y una ventral, superpuestas por su cara basal que secreta una cutícula transparente. Las únicas estructuras visibles en ella son los órganos sensoriales: una queta en cada una de las células y las venas: 5 longitudinales y 2 transversales. Las venas son células epiteliales especializadas presentes en ambas superficies del ala. Durante la diferenciación tardía, las venas dorsales y ventrales se juntan formando un espacio por el que conducen axones y tráqueas (Garcia-Bellido and de Celis, 1992). El proceso de diferenciación de venas ocurre durante los estadios de larva y pupa en los discos imaginales de ala, los tejidos que darán lugar a la futura ala del adulto. Estos tejidos evaginan en la larva L3 y se pliegan de manera que las caras dorsal y ventral quedan de manera adyacente y las

respectivas células contactan por sus membranas basales (Figura 10) (Garcia-Bellido and de Celis, 1992).

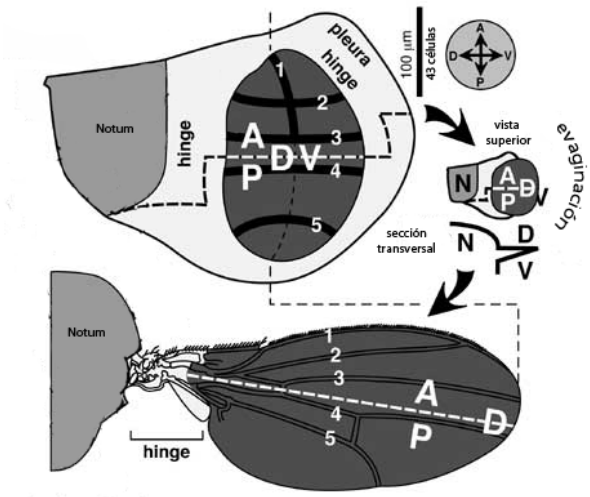


Figura 10. Mapa de destino celular del disco imaginal de ala. Los tonos de gris en el disco imaginal corresponden a los tonos en el ala adulta. Las líneas negras marcan las zonas pre-vena. También se muestra un esquema del proceso de evaginación, donde el disco se pliega para dar lugar al ala adulta. A: anterior; P: posterior; D: dorsal; V: ventral. (Adaptado de InteractiveFly)

La diferenciación de las células de las venas y de la intervena implica una actividad coordinada de diferentes vías de señalización que promueven (EGFR y Dpp) o antagonizan (Notch) la formación de venas (de Celis et al., 1997). En los discos imaginales, la activación de la vía EGFR en filas de células causa su determinación como futuras venas de ala. Consistente con esta función, mutaciones de pérdida de función de los componentes de la vía causan la pérdida de las venas, mientras que alelos de ganancia de función o sobreexpresión de estos componentes resultan en la aparición de tejido venoso ectópico (Diaz-Benjumea and Hafen, 1994).

De manera similar como ocurre en otros tejidos, la especificación de las células de las venas en el disco de ala depende de la señalización EGFR y la regulación negativa de la actividad de Cic (Roch et al., 2002; Blair, 2007; Ajuria et al., 2011). De este modo, la actividad de Cic está restringida a la región intervena, manteniendo la expresión de sus genes diana en la vena, donde Cic está inhibido por la vía (Figura 11A,B) (Roch et al., 2002). Uno de los genes diana que está regulado por la vía EGFR-Cic en este contexto es *argos* (*aos*), el cual codifica para un inhibidor *feedback* de la vía y es requerido para una correcta especificación de las venas (Figura 11A) (Freeman et al., 1992; Roch et al., 2002; Shilo, 2005; Ajuria et al., 2011). De acuerdo con este modelo, pérdidas de función de *cic* causan la formación de tejido venoso ectópico, indicando que Cic restringe la formación de venas a las regiones apropiadas (Figura 11D,E,H,I) (Goff et al.,

2001; Roch et al., 2002), mientras que la sobre-expresión de Cic produce represión en la formación de venas (Figura 11F,G,J) (Ajuria et al., 2011).

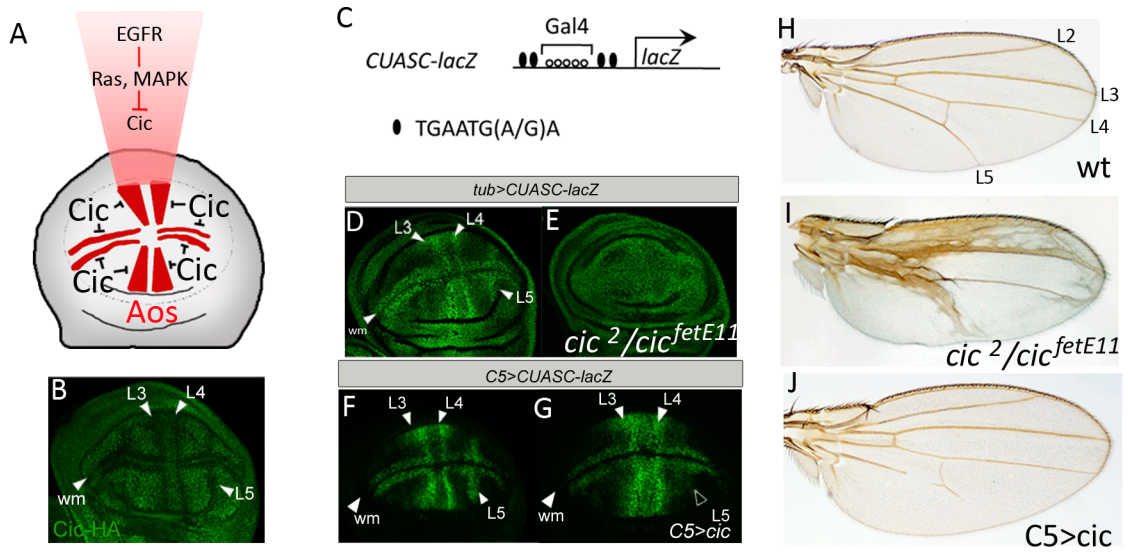


Figura 11. Cic está implicado en la regulación de la venación del ala. (A) Esquema general del papel de Cic en la regulación del patrón del ala en el disco imaginal. La vía EGFR inhibe la acción de Cic, induciendo la expresión de sus genes diana como *aos*. **(B)** Inmunotinción detectando HA un disco imaginal de ala portador de un transgén Cic-HA, donde se aprecia la regulación de los niveles de Cic en respuesta a la vía. **(C)** Esquema del transgén *CUASC-lacZ,* un transgén reportero dirigido por lugares de unión a Gal4[1] flanqueado por lugares de unión a Cic. Este gen reportero reproduce la expresión de los genes diana de Cic. **(D,E)** Inmunotinción detectando lacZ en discos imaginales de ala portadores del transgén reportero *CUASC-lacZ* en un disco *wt* (D) y en un disco mutante *cic* donde su expresión está desreprimida (E). **(F,G)** Inmunotinción detectando lacZ en discos imaginales de ala portadores del transgén reportero *CUASC-lacZ* (F) y expresando UAS-Cic en el disco imaginal, dónde se observa una ganancia de función de Cic al reprimir la expresión del reportero en la vena L5 (ver cabeza de flecha vacía) (G). **(H-J)** Alas adultas de moscas *wt* (H), moscas mutantes para *cic* que muestran tejido venoso ectópico (I) y moscas donde se ha sobre-expresado Cic en el ala, las cuales muestran pérdida de parte de tejido venoso (J). (Adaptado de Ajuria et al., 2011).

3.4. Papel de Cic en proliferación celular

Además de su importante papel en el establecimiento de patrón corporal, Cic también actúa por debajo de la vía EGFR para regular la proliferación de células madre intestinales (ISC) y en discos imaginales larvarios (Tseng et al., 2007; Jiang et al., 2011; Krivy et al., 2013; Jin et al., 2015). En discos de ojo y de ala, mutantes de *cic* incrementan la tasa de proliferación sin afectar el tamaño de la célula y haciendo

[1] El sitema UAS-Gal4 es una herramienta comúnmente utilizada en *Drosophila*, donde el promotor UAS (Upstream Activator Sequence) es activado por el activador transcripcional de levadura Gal4. De este modo, se puede expresar Gal4 en tejidos concretos y proceder a la activación de los genes que llevan como promotor UAS únicamente dónde se está expresando Gal4 (Brand and Perrimon 1993).

innecesaria la señalización EGFR para promover el crecimiento (Tseng et al., 2007). De una manera similar, la inactivación de *cic* induce la proliferación de células madre intestinales en el intestino adulto de *Drosophila* del mismo modo que lo hace la activación ectópica de la señalización EGFR (Jiang et al., 2011; Jin et al., 2015), lo que apoya el modelo de una relación lineal de la vía Ras-MAPK-Cic en la inducción de genes implicados en proliferación.

4. Papel de CIC en enfermedades

Tal como ocurre en *Drosophila*, en mamíferos CIC está regulado por vías RTK para controlar diferentes procesos esenciales (Lee et al., 2002; Dissanayake et al., 2011; Fryer et al., 2011; Lee et al., 2011; Okimoto et al., 2016). El motivo C2, lugar de reconocimiento por MAPK en *Drosophila*, está poco conservado entre *Drosophila* y humanos (Astigarraga et al., 2007) y no se conoce la relevancia funcional de este dominio en humanos. Experimentos bioquímicos han sugerido que CIC se une al mismo lugar de MAPK en ambas especies, pero esta interacción ocurre a través de un motivo distinto de CIC, aunque no hay experimentos funcionales que confirmen este resultado (Figura 2) (Futran et al., 2015). Además, se ha visto que la fosforilación por MAPK y p90RSK regula negativamente a CIC mediante dos mecanismos: por un lado, la fosforilación ocurre en una posición cercana a la HMG-box lo que promueve la unión de CIC a proteínas 14-3-3 (Figura 2), con la consiguiente disminución en la unión de CIC a ADN y causando una desrepresión de sus genes diana. En segundo lugar, la fosforilación de CIC previene el importe nuclear al impedir la unión con KPNA3 (Dissanayake et al., 2011).

CIC se expresa en las células granulares del cerebelo, el hipocampo y el bulbo olfatorio durante el desarrollo del sistema nervioso central, y controla procesos esenciales como la alevolización pulmonar y la homeostasis del ácido biliar (Lee et al., 2002; Dissanayake et al., 2011; Fryer et al., 2011; Lee et al., 2011; Kim et al., 2015). Estudios en células en cultivo y en ratón han demostrado el papel de CIC en la regulación transcripcional del grupo de factores activadores de la transcripción de la familia ETS *ETV/PEA3* (Kawamura-Saito et al., 2006; Dissanayake et al., 2011; Fryer et al., 2011;

Lee et al., 2011). Además, se ha asociado una función anómala de CIC a diferentes enfermedades humanas.

En primer lugar, se ha visto que CIC forma complejos nucleares con ATXN1 y que esta interacción está involucrada en el desarrollo de ataxia espinocerebelar de tipo 1 (Lam et al., 2006; Zoghbi and Orr, 2009). La ataxia espinocerebelar es una enfermedad neurodegenerativa causada por la expansión de la región poliglutamínica (poliQ) de ATXN1 (Orr et al., 1993). Tal como se ha comentado en apartados anteriores, en mamíferos CIC requiere del correpresor ATXN1 para ejercer su función represora (Lam et al., 2006; Bowman et al., 2007; Lim et al., 2008; Crespo-Barreto et al., 2010; Lee et al., 2011), de modo que las formas de ATXN1 con la región poliQ extendida causan una función represora anómala de CIC (Lam et al., 2006; Lim et al., 2008; Fryer et al., 2011). El aumento de las repeticiones CAG pueden causar, por un lado, que la interacción de CIC con ATXN1 se debilite, provocando una represión insuficiente de sus genes diana *ETS* como *ETV5* (Lim et al., 2008; Fryer et al., 2011) o, por otro lado, puede causar efectos totalmente contrarios, provocando una unión más fuerte de CIC a otros genes diana, provocando la hiperrepresión, siendo este efecto regulado por la vía EGFR (Fryer et al., 2011).

En segundo lugar, se ha visto que CIC está implicado en la aparición de diferentes tipos de cáncer como cáncer de mama, colon, oligodendroglioma y también de metástasis. En cáncer CIC se comporta como un supresor tumoral y de metástasis, de manera que mutaciones que inactivan la proteína pueden llevar a la aparición de estas patologías (Sjöblom et al., 2006; Bettegowda et al., 2011; Seshagiri et al., 2012; Okimoto et al., 2016). El patrón de mutaciones inactivantes de *CIC* es particularmente interesante en oligodendroglioma, donde se ha visto que este tipo de mutaciones somáticas afectan de manera especialmente recurrente a dos regiones de la proteína: el dominio HMG-box y el motivo C1 (Figura 12) (Bettegowda et al., 2011; Jiao et al., 2012; Sahm et al., 2012; Yip et al., 2012; Chan et al., 2014; Chittaranjan et al., 2014; Gleize et al., 2015; Padul et al., 2015).

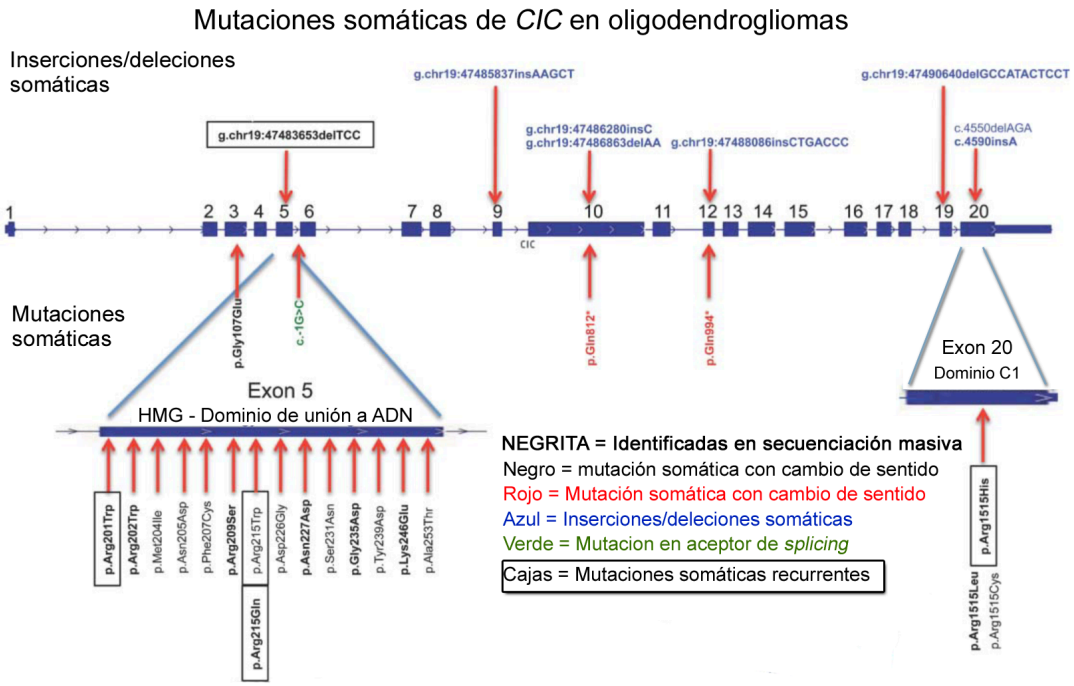


Figura 12. Mutaciones somáticas de *CIC* en oligodendroglioma. Las mutaciones están representadas a lo largo de toda la proteína CIC, donde la HMG-box y el motivo C1 están indicados. Las mutaciones somáticas están representadas con flechas y el cambio de aminoácido. Las mutaciones recurrentes están indicadas dentro de cajas. La mayoría de mutaciones están concentradas entre los exones 5 y 20, coincidiendo con los dominios HMG-box y C1. (Adaptado de Yip et al., 2012)

Además, se ha visto que CIC está implicado en la aparición de un tipo de sarcomas de tejido blando llamados sarcomas de tipo *Ewing-like* (Kawamura-Saito et al., 2006). En este tipo de sarcomas se dan translocaciones cromosómicas entre los cromosomas 19 y 4 o 10, dando como resultado una fusión entre las proteínas CIC y el factor de transcripción Double Homeobox 4 (DUX4). Estas fusiones suelen conservar casi toda la región codificante de la proteína CIC (incluido el motivo C1) y solo la fracción final de la proteína DUX4 (Kawamura-Saito et al., 2006; Yoshimoto et al., 2009; Graham et al., 2012; Italiano et al., 2012; Choi et al., 2013; Machado et al., 2013) y esta fusión provoca un cambio en la actividad de CIC, dónde éste se convierte en un activador transcripcional en lugar de un represor, regulando positivamente los genes diana de la familia *ETV/PEA3*, y dando lugar a la aparición de sarcomas (Kawamura-Saito et al., 2006).

Por último, se han encontrado mutaciones en *CIC* como causantes de resistencias a tratamientos para inhibir la señal EGFR en células en cultivo de cáncer de pulmón de células no pequeñas y de páncreas (Liao et al., 2017; Wang et al., 2017). Estos tipos de cáncer a menudo son causados por mutaciones que promueven la sobre-activación de

la vía EGFR (Blasco et al., 2011; Stewart et al., 2015), de modo que son tratados con moléculas inhibidoras de la señal (Lynch et al., 2004; Paez et al., 2004; Wang et al., 2016). A pesar de que las células cancerosas suelen responder bien al tratamiento, no todas las células responden correctamente, incluso a veces aparecen resistencias (Mok et al., 2009). Estas resistencias suelen aparecer como resultado de mutaciones en genes que actúan por debajo de EGFR y que contrarrestan el efecto del tratamiento (Sharifnia et al., 2014). En *screenings* para identificar factores cuyas mutaciones pueden resultar en resistencias a tratamientos con inhibidores de EGFR en células tumorales de pulmón y de páncreas se han encontrado mutaciones en *CIC* que resultan en una expresión ectópica de los genes *ETV* (Liao et al., 2017; Wang et al., 2017). Además, mutaciones en el factor ATXN1L causan una disminución de los niveles de CIC, lo que resulta en el mismo efecto de desrepresión de los genes a los que está regulando (Wang et al., 2017).

Estas evidencias demuestran la importancia biomédica de estudiar los mecanismos por los que CIC ejerce su función, para poder entender mejor la patogénesis de las enfermedades humanas en las que está implicado y ayudar a diseñar nuevos tratamientos.

OBJETIVOS

Durante los últimos años, estudios genéticos y moleculares han mostrado la relevancia del papel de Cic como efector de las vías RTK tanto en *Drosophila* como en humanos. Se ha visto que en *Drosophila* Cic reprime sus genes diana por debajo de las vías de Torso y EGFR, y que en algunos casos lo está haciendo con la ayuda del correpresor Gro. Hasta el momento se desconoce si la dependencia de Gro esta conservada en otras especies.

Otro aspecto que se desconoce de Cic es si tiene mecanismos adicionales a la regulación por la vía RTK para regular la estabilidad y función de Cic en *Drosophila*.

Por otro lado, otra característica que se ha visto esencial en la función represora de Cic es la presencia del motivo C1, imprescindible para la actividad de la proteína. En humanos, tal como ocurre en *Drosophila*, CIC ejerce su función por debajo de la vía de EGFR para regular funciones esenciales en el desarrollo y actuar como supresor tumoral en la mayoría de tipos de cáncer y como oncogén en los casos de sarcomas de tipo *Ewing-like*. En todos estos contextos el motivo C1 parece esencial para la regulación de sus funciones.

Así pues, durante esta tesis se ha intentado resolver algunas de las cuestiones abiertas sobre el mecanismo de acción de Cic y se ha intentado caracterizar el mecanismo molecular por el que Cic reprime a sus genes diana, siendo los principales objetivos:

1. Estudiar la relación funcional entre Cic y Gro en el embrión temprano y otros tejidos de *Drosophila* y averiguar si esta relación está conservada en otras especies.
2. Estudiar la sensibilidad de Cic a la vía a través del motivo C2 e investigar si hay mecanismos de regulación adicionales a las vías RTK en *Drosophila*.
3. Caracterizar la función del motivo C1 para entender el mecanismo molecular por el que es necesario para las funciones represoras de la proteína en *Drosophila*, y su relación funcional con diferentes tipos de cáncer en humanos.

PUBLICACIONES

INFORME SOBRE LA CONTRIBUCIÓN DE LA DOCTORANDA A LAS PUBLICACIONES DE ESTA TESIS DOCTORAL

La memoria de la tesis doctoral de **Marta Forés Maresma**, titulada **Mecanismo de acción del factor represor Capicua, un sensor de señales Ras/MAPK**, se presenta como un compendio de las cuatro siguientes publicaciones:

ARTÍCULO 1:

Título: Origins of context-dependent gene repression by Capicua

Autores: Marta Forés, Leiore Ajuria, Núria Samper, Sergio Astigarraga, Claudia Nieva, Rona Grossman, Sergio González-Crespo, Ze'ev Paroush, Gerardo Jiménez

Referencia: *PLoS Genet* **11**, e1004902 (2015)

Este artículo presenta un extenso trabajo experimental que ha sido realizado en su mayor parte por Marta Forés y Leiore Ajuria. El artículo describe las actividades reguladoras de Capicua (Cic) en varios contextos del desarrollo, y su relación con el correpresor Groucho en el embrión temprano. También explora el origen evolutivo y diversificación de distintas isoformas de Cic, describiendo la aparición de un nuevo dominio con funciones esenciales en la formación del patrón corporal de *Drosophila*. Estos resultados son interesantes porque ilustran uno de los pocos ejemplos conocidos acerca del origen y especialización de una función molecular esencial para el desarrollo. Para llevar a cabo este trabajo, Marta ha generado y analizado las líneas transgénicas presentadas en las Figuras 4, 5 y 6 del artículo. Además, su trabajo ha supuesto un extenso conjunto de experimentos genéticos en distintos fondos mutantes, algunos de ellos muy elaborados. Finalmente, Marta ha realizado la mayoría de los análisis embriológicos e histoquímicos, así como los correspondientes análisis de microscopía y procesamiento de imágenes de las Figuras 3, 4, 5, 6 y S1.

En resumen, se trata de un estudio complejo y muy exigente técnicamente que Marta ha desarrollado de manera impecable gracias a su iniciativa, creatividad y rigor técnico.

ARTÍCULO 2:

Título: Using CRISPR-Cas9 to study ERK signaling in *Drosophila*

Autores: Marta Forés, Aikaterini Papagianni, Laura Rodríguez-Muñoz, Gerardo Jiménez

Referencia: In *ERK Signaling: Methods and Protocols* (ed. G. Jiménez), pp. 353-365. New York, NY: Springer New York (2017)

Este trabajo presenta un método para analizar la señalización por Ras-MAPK en *Drosophila* mediante la tecnología de CRISPR-Cas9. En concreto, se describe el uso de la técnica para generar una mutación en un motivo de interacción (*docking site*) con MAPK de la proteína Cic. El trabajo tiene además el interés de ilustrar la obtención de una mutación de ganancia de función *in vivo*, describiendo los conceptos y dificultades técnicas del proceso. Para la realización de este trabajo, Marta ha analizado 50 mutaciones en *cic* hasta encontrar un nuevo alelo con las características genéticas deseadas, al que hemos denominado cic^{3}. Además, Marta ha caracterizado las ganancias de función asociadas a esta mutación en distintos procesos del desarrollo. En resumen, el trabajo ha supuesto un esfuerzo considerable que Marta ha sabido completar prácticamente sin necesitar ayuda hasta obtener una mutación única en un regulador de gran interés.

ARTÍCULO 3:

Título: Minibrain and Wings apart control organ growth and tissue patterning through down-regulation of Capicua

Autores: Liu Yang, Sayantanee Paul, Kenneth G. Trieu, Lucas G. Dent, Francesca Froldi, **Marta Forés**, Kaitlyn Webster, Kellee R. Siegfried, Shu Kondo, Kieran Harvey, Louise Cheng, Gerardo Jiménez, Stanislav Y. Shvartsman, Alexey Veraksa

Referencia: *Proc Natl Acad Sci USA* **113**, 10583-10588 (2016)

Este trabajo muestra como la quinasa Mnb junto con la proteína adaptadora Wap fosforilan a Cic para inactivar sus funciones como regulador del crecimiento de diferentes tejidos y la formación del patrón de venas del ala. Además, se muestra que durante la regulación del patrón del ala MAPK y Mnb/Wap tienen funciones aditivas en la inactivación de Cic. Este trabajo abre la posibilidad de una función de Cic como integrador de distintas señales transmitidas por diferentes vías reguladoras. Durante el trabajo, Marta ha caracterizado la función aditiva de las vías MAPK y Mnb/Wap

utilizando el alelo de *cic* que ella misma generó en la publicación anterior. En concreto, Marta ha generado adultos recombinantes que combinan el alelo cic^3 y la expresión dirigida en el ala de un ARN de interferencia contra Mnb, analizando los fenotipos resultantes que se muestran en la Figura 6 del artículo.

ARTÍCULO 4:

Título: A new mode of DNA binding distinguishes Capicua from other HMG-box factors and explains its mutation patterns in cancer

Autores: Marta Forés, Lucía Simón-Carrasco, Leiore Ajuria, Núria Samper, Sergio González-Crespo, Matthias Drosten, Mariano Barbacid, Gerardo Jiménez

Referencia: *PLoS Genet* **13**, e1006622 (2017)

En este trabajo se describe el mecanismo de unión a ADN de Cic. Desde su descubrimiento hace más de 15 años, se ha asumido que Cic reconoce al ADN mediante su dominio HMG-box. En cambio, Marta ha demostrado que dicho reconocimiento requiere también el dominio denominado C1, cuya función era desconocida. El dominio C1 se encuentra ampliamente conservado en la evolución y resulta frecuentemente inactivado por mutaciones en tumores humanos. Por tanto, el trabajo ha identificado un mecanismo de gran relevancia para la función y evolución de las proteínas HMG-box, y ha definido un aspecto clave de la función de Cic como supresor tumoral en humanos. Durante este trabajo, Marta ha generado mediante la técnica de CRISPR-Cas9 un nuevo alelo de *cic* que hemos denominado cic^4. Este alelo contiene una mutación en el dominio C1, y Marta ha caracterizado los efectos que produce en distintos contextos del desarrollo. También ha generado las construcciones mutantes en el motivo C1 (ΔC1 y R1515L) para los ensayos en células mostrados en la Figura 4. Además, ha generado todas las construcciones probadas en los ensayos de EMSA, expresando las proteínas correspondientes *in vitro* o en bacteria y analizándolas a continuación en este ensayo. Este trabajo ha sido especialmente laborioso, ya que además de las construcciones mostradas en el artículo, Marta ha diseñado y analizado la unión a ADN de otras muchas proteínas, sumando en total más de 50 construcciones. También ha analizado las secuencias de diferentes *enhancers* reconocidos por Cic y ha obtenido un transgén sintético para analizar el efecto de las

secuencias adyacentes al motivo de unión de Cic *in vivo.* Finalmente, Marta ha generado para este trabajo el primer modelo *in vivo* de la actividad de la oncoproteína quimera CIC-DUX4. Para ello, ha generado líneas transgénicas que expresan dicha proteína en *Drosophila*, las ha editado posteriormente mediante CRISPR-Cas9 y las ha analizado molecularmente utilizando el modelo del disco imaginal de ala. En conjunto, Marta ha realizado por completo los experimentos y la adquisición de las imágenes de las Figuras 2, 5, 6 y 7. Se trata pues de un trabajo de gran envergadura técnica y conceptual, y con amplias implicaciones tanto desde el punto de vista básico como clínico. De hecho, el trabajo ha merecido la publicación de una nota de prensa por parte del departamento de comunicación del CSIC, la cual ha sido recogida en publicaciones de información biomédica y en la prensa generalista.

En definitiva, Marta ha realizado un trabajo de gran calidad y que cumple con creces el nivel exigido por cualquier universidad internacional. La relevancia de su contribuciones científicas está avalada no solo por las publicaciones presentadas, sino también por la calidad y alcance de sus resultados en áreas complementarias como la biología molecular, la biología evolutiva, la genética y la biomedicina. En mi opinión, su trabajo merece la mejor consideración por parte de la comisión evaluadora.

Gerardo Jiménez

INFORME SOBRE EL FACTOR DE IMPACTO DE LAS PUBLICACIONES

Las publicaciones presentadas en esta tesis presentan resultados de gran relevancia y que suponen un avance de interés general para la comunidad científica. Los cuatro trabajos listados se han publicado en las revistas *PLOS Genetics* y *Proceedings of the National Academy of Sciences of USA* (*PNAS*) y en el libro *ERK Signaling - Methods and Protocols* de la serie *Methods in Molecular Biology*, siendo todos de ámbito internacional. La revista *PLOS Genetics* tiene un elevado prestigio en los campos de la genética, la regulación génica, la genómica y otras áreas relacionadas, mientras que *PNAS* es una de las revista multidisciplinares más citadas en el mundo. Los índices de impacto para estas revistas en 2015 son: 6,661 para *PLOS Genetics* y 9,423 para *PNAS*. Estos índices de impacto sitúan a estas revistas en el primer cuartil de publicaciones en sus categorías. La serie *Methods in Molecular Biology*, aun estando recogida en las bases bibliográficas más importantes, no cuenta con un índice de impacto calculado.

PUBLICACIÓN 1: Origins of context-dependent gene repression by Capicua

Autores:

Marta Forés, Leiore Ajuria, Núria Samper, Sergio Astigarraga, Claudia Nieva, Rona Grossman, Sergio González-Crespo, Ze'ev Paroush, Gerardo Jiménez

Referencia:

PLoS Genet **11**, e1004902 (2015)

Resumen:

Las vías de señalización RTK inducen múltiples respuestas biológicas, a menudo regulando la expresión de genes que están por debajo de la vía. La proteína HMG-box Capicua (Cic) es un represor transcripcional que esta inhibido en respuesta a la señal RTK, permitiendo la inducción de los genes diana de Cic de manera dependiente de la vía. Tanto en *Drosophila* como en mamíferos, Cic se expresa como dos isoformas, una larga (Cic-L) y una corta (Cic-S), de las cuales su significancia funcional y su mecanismo de acción no se conocen del todo bien. En este artículo mostramos que la proteína Cic de *Drosophila* necesita el correpresor Groucho (Gro) en su función en el embrión temprano, pero no durante otros estadios del desarrollo. Este mecanismo dependiente de Gro requiere un pequeño motivo peptídico, único en la isoforma Cic-S y llamado N2, el cual es distinto a los motivos de interacción con Gro definidos hasta ahora, y funciona como un elemento represor autónomo y transferible. Inesperadamente, nuestros datos indican que el motivo N2 es una innovación evolutiva que se originó en los insectos dípteros, cuando la isoforma Cic-S evolucionó de la isoforma Cic-L. De acuerdo con esto, la isoforma Cic-L carece del motivo N2 y es completamente inactiva en el embrión temprano de *Drosophila*, indicando que el motivo N2 aportó a la forma Cic-S una nueva actividad dependiente de Gro en este estadio. Sugerimos que las funciones co-reguladoras de Cic-S y Gro han facilitado la evolución de una compleja red transcripcional por la señalización de Torso en las moscas modernas. Notablemente, nuestros resultados implican que las proteínas Cic de mamíferos improbablemente actúan a través de Gro, ya que su forma Cic-S se

debe haber originado independientemente de la de *Drosophila*. Por lo tanto, las proteínas Cic emplean mecanismos represores distintos que están asociados a discretos cambios estructurales en la historia evolutiva de esta familia de proteínas.

OPEN ACCESS Freely available online

PLOS | GENETICS

Origins of Context-Dependent Gene Repression by Capicua

Marta Forés[1☯], Leiore Ajuria[1☯], Núria Samper[1], Sergio Astigarraga[1¤], Claudia Nieva[1], Rona Grossman[2], Sergio González-Crespo[1], Ze'ev Paroush[2], Gerardo Jiménez[1,3*]

1 Institut de Biologia Molecular de Barcelona-CSIC, Parc Científic de Barcelona, Barcelona, Spain, **2** Department of Developmental Biology and Cancer Research, IMRIC, The Hebrew University, Jerusalem, Israel, **3** Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

Abstract

Receptor Tyrosine Kinase (RTK) signaling pathways induce multiple biological responses, often by regulating the expression of downstream genes. The HMG-box protein Capicua (Cic) is a transcriptional repressor that is downregulated in response to RTK signaling, thereby enabling RTK-dependent induction of Cic targets. In both *Drosophila* and mammals, Cic is expressed as two isoforms, long (Cic-L) and short (Cic-S), whose functional significance and mechanism of action are not well understood. Here we show that *Drosophila* Cic relies on the Groucho (Gro) corepressor during its function in the early embryo, but not during other stages of development. This Gro-dependent mechanism requires a short peptide motif, unique to Cic-S and designated N2, which is distinct from other previously defined Gro-interacting motifs and functions as an autonomous, transferable repressor element. Unexpectedly, our data indicate that the N2 motif is an evolutionary innovation that originated within dipteran insects, as the Cic-S isoform evolved from an ancestral Cic-L-type form. Accordingly, the Cic-L isoform lacking the N2 motif is completely inactive in early *Drosophila* embryos, indicating that the N2 motif endowed Cic-S with a novel Gro-dependent activity that is obligatory at this stage. We suggest that Cic-S and Gro coregulatory functions have facilitated the evolution of the complex transcriptional network regulated by Torso RTK signaling in modern fly embryos. Notably, our results also imply that mammalian Cic proteins are unlikely to act via Gro and that their Cic-S isoform must have evolved independently of fly Cic-S. Thus, Cic proteins employ distinct repressor mechanisms that are associated with discrete structural changes in the evolutionary history of this protein family.

Citation: Forés M, Ajuria L, Samper N, Astigarraga S, Nieva C, et al. (2014) Origins of Context-Dependent Gene Repression by Capicua. PLoS Genet 11(1): e1004902. doi:10.1371/journal.pgen.1004902

Editor: Norbert Perrimon, Harvard Medical School, Howard Hughes Medical Institute, United States of America

Received October 23, 2014; **Accepted** November 17, 2014; **Published** January 8, 2015

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by research grants from the Spanish Government (BFU2008-01875 and BFU2011-23611), Generalitat de Catalunya (2009SGR-1075) and Fundació La Marató de TV3 (20131730). GJ is an ICREA investigator. ZP is supported by grants from the National Institute of General Medical Sciences (NIH R01GM086537), the Israel Science Foundation (Center of Excellence; 1772/13) and the Król Charitable Foundation. ZP is an incumbent of The Lady Davis Chair in Experimental Medicine and Cancer Research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: gjcbmc@ibmb.csic.es

☯ These authors contributed equally to this work.

¤ Current address: Skirball Institute of Biomolecular Medicine, New York, New York, United States of America

Introduction

Receptor Tyrosine Kinase (RTK) signaling pathways regulate tissue development and morphogenesis in all metazoans [1]. RTKs often signal through the conserved Ras-Raf-MAPK cascade, leading to phosphorylation of nuclear transcription factors which then elicit changes in target gene expression. The HMG-box protein Capicua (Cic) has recently emerged as a general nuclear sensor of RTK signaling pathways [2]. Originally discovered downstream of the Torso RTK in *Drosophila* embryogenesis, Cic has been subsequently shown to function downstream of other RTKs at multiple stages of fly development [3–11]. In all cases, Cic represses transcription of RTK-responsive genes in unstimulated cells, whereas activation of RTK signaling results in phosphorylation and downregulation of Cic and this causes derepression of its target genes [7,10,12,13].

Cic is highly conserved from cnidarians to vertebrates and is implicated in several human pathologies such as spinocerebellar ataxia type 1 (SCA1) and oligodendroglioma (OD) [14–17]; reviewed in [2]. Indeed, Cic proteins from *Drosophila* and mammals share many functional and structural properties: they repress transcription by binding to related DNA sites in target genes, appear to be similarly downregulated by RTKs and are expressed as two main isoforms, short (Cic-S) and long (Cic-L), which differ in their N-terminal regions [7,9,10,14,15,17–19]. However, despite these similarities, it is currently unclear whether all Cic family proteins employ a common mechanism of repression. Studies in mouse and human cells have shown that Cic associates with Ataxin1 (Atxn1), a co-repressor involved in SCA1 [14,15,17,20,21]. On the other hand, previous studies in *Drosophila* have suggested that Cic functions together with Groucho (Gro) [3,10], a WD-repeat co-repressor that associates with multiple repressors, including Hairy/Hes, Nkx, Lef/Tcf and Runx family proteins (reviewed in [22,23]). However, the functional links between Cic and Gro remain unclear, since no

Author Summary

Understanding the evolution of developmental regulatory mechanisms is a central challenge of biology. Here we uncover a newly evolved mechanism of transcriptional repression by Capicua (Cic), a conserved sensor of Receptor Tyrosine Kinase (RTK) signaling. In *Drosophila*, Cic patterns the central regions of the embryo by repressing genes induced by Torso RTK signaling at the poles. We show that Cic performs this function by recruiting the Groucho (Gro) corepressor and that this mechanism is an evolutionary innovation of dipteran insects. Indeed, we find that recruitment of Gro depends on a short motif of Cic (N2) specific to dipterans. Strikingly, moreover, the form of Cic that existed before the origin of dipterans is completely inactive in fly embryos, whereas the equivalent form carrying N2 displays significant function. This suggests that evolution of the N2 motif caused a fundamental change in Cic repressor activity, which we propose has enabled the complex roles of Cic, Gro and Torso signaling in fly embryonic patterning. In contrast, Cic functions independently of Gro in other *Drosophila* tissues and probably also in mammals, where Cic lacks the N2 sequence. Thus, our results illustrate the structural and evolutionary origins of essential functional variations within a highly conserved family of developmental regulators.

molecular interaction between these proteins has been validated in vivo [2,24].

Here, we investigate the mechanism of *Drosophila* Cic repression and its relationship with Gro. We find that Cic functions via Gro in the early embryo but not at other developmental stages. The Gro-assisted mechanism depends on a previously unrecognized motif of Cic (N2), which is essential for recruitment of Gro in vivo. Remarkably, the N2 motif is highly conserved among Cic orthologues in flies and mosquitoes, but is absent in all other species, suggesting that it originated in ancestral dipterans. Furthermore, the N2 domain appears to be a structural innovation associated with the emergence of fly Cic-S isoforms from a pre-existing Cic-L-like isoform. This implies that mammalian Cic proteins, which lack the N2 motif, probably function independently of Gro, and that their Cic-S isoforms must have evolved independently of fly Cic-S. Thus, Cic proteins exhibit context-dependent repressor activities that are partly associated with key structural changes that have occurred during the evolution of this protein family.

Results

Context-dependent activities of Cic in *Drosophila* development

Cic and Gro are both essential for repression of two terminal gap genes, *tailless* (*tll*) and *huckebein* (*hkb*), in central regions of the blastoderm embryo; this repression is normally relieved by Torso RTK signaling at the embryonic termini, thereby enabling localized induction of *tll* and *hkb* by broadly distributed activators [3,4,25]. These shared requirements of Cic and Gro in the terminal system have led to the idea that both proteins act in a common repressor complex (see refs. [2,24]). However, we have assayed the requirement of Gro for Cic repressor functions in other developmental contexts and found that Gro is dispensable for such functions (Fig. 1). Specifically, we examined two systems - the developing wing and the ovarian follicular epithelium- where Cic represses specific target genes such as *argos* and *mirror*, respectively, under the control of the EGFR pathway [4–6,10–12]. In these experiments, we compared the effects caused by the loss of Cic or Gro function using mosaic analyses. Unexpectedly, we found that loss of Gro function does not impair Cic repression in any of those systems, indicating that Cic represses *argos* and *mirror* independently of Gro (Fig. 1).

In light of these results, we have re-evaluated the functional links between Cic and Gro in the early embryo. First, we asked if Cic-mediated repression of a synthetic reporter gene relies on Gro activity in the early embryo. To this end, we used a transgenic construct containing a minimal *hunchback* (*hb*) enhancer linked to a pair of individual Cic binding sites (*hbC*; ref. [10]) (Fig. 2A). The intact *hb* enhancer drives broad expression in the anterior third of the embryo (Fig. 2B), whereas *hbC* is repressed by Cic and drives expression only in the anterior pole of the embryo, where Cic is downregulated by Torso RTK signaling (Fig. 2C, D). As shown in Fig. 2E, we find clear derepression of *hbC* activity in embryos lacking Gro function, implying that Cic represses *hbC* via Gro in this assay. These results support the idea that Cic indeed acts through Gro in early embryonic patterning.

N2, a new motif of Cic that is essential for repression

Cic does not contain either of the two previously defined Gro-binding motifs present in known Gro-dependent repressors, the WRPW- and eh1-like peptides [22], and we have not detected direct interactions between functionally important regions of Cic and Gro [10]. Therefore, we asked what sequences of Cic mediate its Gro-dependent repressor activity. Assuming that those sequences could be evolutionarily conserved, we noted a novel conserved motif present at the N-terminus of the Cic-S isoform (GenBank protein AAF55751), which we designate N2 (Fig. 3A, B). This motif is encoded in two adjacent exons: a 5′ exon specific of the *cic-S* transcript and a 3′ exon shared by both *cic-S* and *cic-L* transcripts (see also below). The sequence encoded by the *cic-S*-specific exon (LYLQCLL) is conserved in dipteran species (Fig. 3A, B, highlighted in red), whereas the peptide common to Cic-S and Cic-L isoforms (SLSSSRSATP) is conserved from hydra to humans (Fig. 3A, B, highlighted in black). To assess the functional significance of N2, we assayed the activity of a Cic-S derivative lacking this motif ($Cic^{\Delta N2}$). We find that $Cic^{\Delta N2}$ is expressed at normal levels in transgenic embryos but is unable to repress *tll*, a *tll* reporter or *hkb* (Fig. 3C-J). Accordingly, $Cic^{\Delta N2}$ does not provide any rescue of the *cic* embryonic mutant phenotype (Fig. 3K-M), indicating that N2 is critical for Cic function in the early embryo.

We also tested two mutations affecting each of the sub-elements of N2. Surprisingly, disruption of the Cic-S-specific element caused a complete loss of Cic-S function, whereas mutation of the second, highly conserved sequence had a minor effect on protein activity (Fig. 3N, O). Thus, only the dipteran-specific portion of N2 is essential for Cic embryonic function.

N2 is a Gro-dependent repressor element

Based on the above results, we hypothesized that N2 could be involved in recruiting Gro to Cic target genes. In fact, the critical N2 sequence shares some similarity with the consensus eh1 motif (FxIxxIL) that binds directly to Gro, although it lacks the characteristic phenylalanine residue at position 1. We therefore tested if N2 functions as an autonomous, transferable Gro-dependent motif capable of imposing repressor activity on a heterologous DNA-binding domain. For this, we adopted the *Sex-lethal* (*Sxl*) repression assay, an in vivo strategy for analyzing the activity of known or potential repressor domains [26,27]. In this assay, a domain under analysis is used to replace the Gro-binding

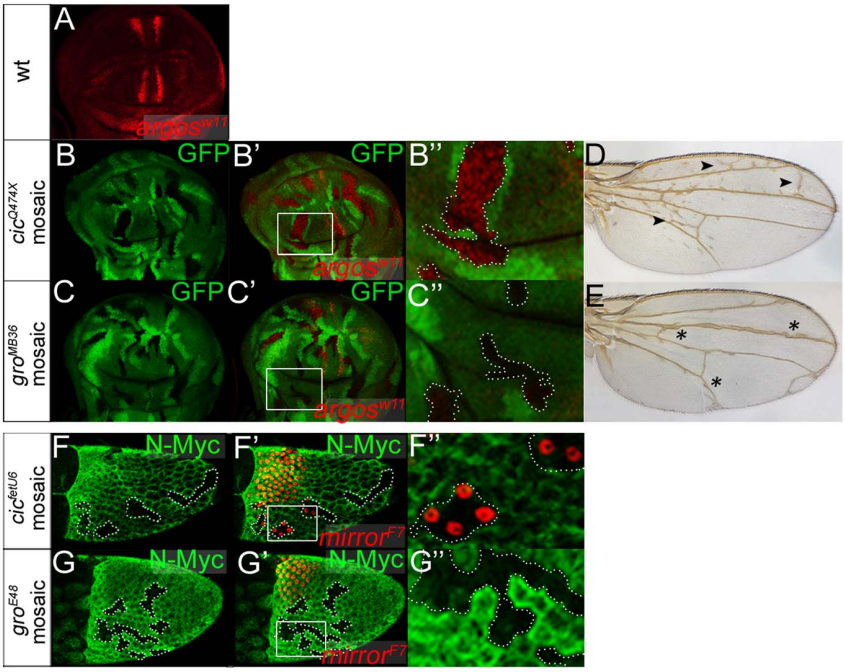


Fig. 1. Cic functions independently of Gro in the ovary and in the wing. (A) Expression of *argos* in a third instar wing imaginal disc as revealed by LacZ (β-galactosidase) immunostaining using the $argos^{W11}$*–lacZ* enhancer trap. Expression is detected in presumptive vein stripes where EGFR signaling is active, and is absent in intervein regions where Cic represses *argos*. (B-C″) Mosaic wing imaginal discs carrying cic^{Q474X} (B-B″) and gro^{MB36} (C-C″) mutant clones marked by absence of GFP (green, outlined in B″ and C″). B′ and C′ show merged images of GFP signals and $argos^{W11}$*–lacZ* expression (red); B″ and C″ show close-ups of boxed areas in panels B′ and C′. Note that loss of Cic function leads to full derepression of $argos^{W11}$*–lacZ* in the mutant clones, whereas the loss of Gro causes derepression of $argos^{W11}$*–lacZ* only in close proximity to its normal stripes of expression. This localized effect of Gro probably reflects its role together with Enhancer-of-split/Hes repressors in refining *argos* expression [45]. (D and E) Mosaic adult wings carrying cic^{Q474X} (D) and gro^{MB36} (E) mutant clones induced in third instar larvae as above. Consistent with the effects on $argos^{W11}$*–lacZ* expression, the phenotypes of cic^{Q474X} and gro^{MB36} mosaic wings are clearly different: *cic* mosaic wings show patches of ectopic vein material throughout the wing blade (arrowheads), whereas *gro* mosaic wings display localized thickening of veins (asterisks). This indicates that Cic repression in the developing wing does not rely on Gro. (F-G″) Stage-10 mosaic egg chambers carrying cic^{fetU6} (F-F″) and gro^{E48} (G-G″) mutant clones marked by absence of N-Myc immunofluorescence (green, outlined in F″ and G″). F′ and G′ show merged images of N-Myc signals and *mirror* expression visualized using the $mirror^{F7}$*–lacZ* enhancer trap and anti-LacZ staining (red). F″ and G″ show close-ups of boxed areas in panels F′ and G′. *mirror* is a key regulator of dorsoventral axis formation that is activated by EGFR signaling in dorsal-anterior follicle cells, and repressed by Cic in ventral follicle cells. *cic* loss-of-function clones in ventral regions cause derepression of $mirror^{F7}$*–lacZ*, although only in the anterior half of the follicular epithelium [4,6]. In contrast, *gro* mutant clones do not show $mirror^{F7}$*–lacZ* derepression, suggesting that Cic also acts independently of Gro in this context.

doi:10.1371/journal.pgen.1004902.g001

WRPW motif of the Hairy repressor and tested for its ability to repress *Sxl* expression in the embryo (Fig. 4A). Using this approach, we found that a Hairy chimera carrying the N2 motif instead of the WRPW peptide ($Hairy^{N2}$) represses *Sxl* as efficiently as intact Hairy (Fig. 4B-D). In contrast, four control Hairy chimeras carrying a mutant version of N2 or other conserved motifs from Cic, did not (Fig. 4B, E-H; see also S1 Fig.). Moreover, repression by the $Hairy^{N2}$ chimera depends on Gro, as it is lost in gro^{E48} mutant embryos that lack Gro activity (Fig. 4I). This indicates that N2 is a discrete, Gro-dependent repressor motif.

We also analyzed the activity of $Hairy^{N2}$ in the presence of a Gro mutant protein, Gro^{MB41}, which can not bind to WRPW or eh1 motifs but retains normal function in the terminal system (potentially acting together with Cic) [28]. The Gro^{MB41} mutant carries an amino acid substitution (R483H) affecting the central pore of the Gro β propeller domain, thereby preventing binding of WRPW or eh1 motifs across this pore [28]. We found that $Hairy^{N2}$ displays significant repressor activity in gro^{MB41} embryos, whereas native Hairy is completely inactive in this background (Fig. 4J, K). Thus, Gro^{MB41} is functional both in repressing terminal gap genes and in mediating repression by $Hairy^{N2}$, suggesting that it is recruited in each of these systems through similar interactions that involve the N2 motif.

As an independent test of this idea, we analyzed a Cic derivative in which the N2 sequence was replaced by the eh1 motif (FSISNIL) from the Engrailed homeodomain protein (Cic^{eh1}; Fig. 5A). If Gro is recruited to the terminal system through the N2 motif, replacing this motif by the eh1 element should render Gro^{MB41} non-functional in that system. For these experiments, we monitored the expression of the central gap gene *knirps* (*kni*) as a sensitive readout of Cic and Cic^{eh1} repressor activities (Fig. 5B). *kni* is a target of the Tll repressor. When Cic is active, it restricts *tll* expression to the posterior pole of the embryo, thereby permitting expression of *kni* in the presumptive abdomen (Fig. 5B, C). In contrast, loss of Cic function causes derepression of *tll* and corresponding loss of the central *kni* stripe (Fig. 5D, E). We find that Cic^{eh1} is an active repressor capable of rescuing *kni* expression in *cic* mutant embryos (Fig. 5F), indicating that the eh1 peptide can compensate for the loss of endogenous N2 in its normal setting. We then compared *kni* expression in gro^{MB41} embryos expressing either endogenous Cic or Cic^{eh1}. As previously

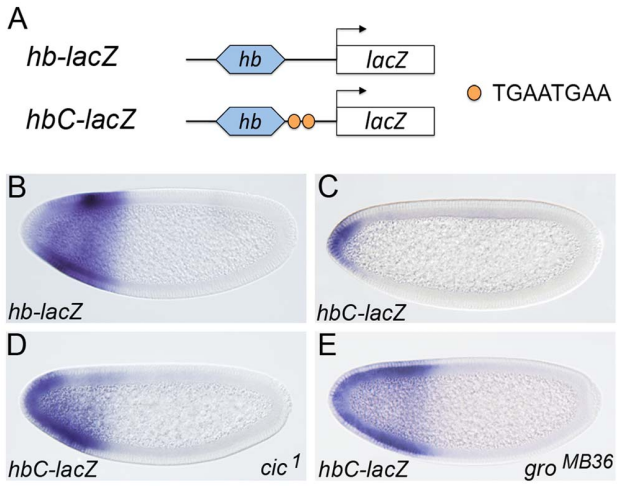


Fig. 2. Cic binding sites are sufficient for recruitment of Gro in vivo. (A) Diagram of *lacZ* transgenes under the control of a minimal *hb* enhancer and canonical Cic binding sites (TGAATGAA). (B-E) mRNA expression patterns of *hb-lacZ* and *hbC-lacZ* in otherwise wild-type (B, C), cic^{1} (D) or gro^{MB36} (E) mutant embryos; note the strong derepression of *hbC-lacZ* in both mutant backgrounds. In this and subsequent figures, anterior is to the left and dorsal is up.
doi:10.1371/journal.pgen.1004902.g002

reported, *kni* expression is normal in the first case (ref. [28]; Fig. 5G), whereas there is clear loss of *kni* expression in the presence of Cic^{ch1} (Fig. 5H). Therefore, it is the presence of an intact N2 motif in Cic that enables Gro^{MB41} to be functional in the terminal system, supporting our conclusion that N2 links Cic and Gro in the *Drosophila* embryo.

Origin of N2 and Cic-S in dipterans

As indicated above, the key repressor element within the N2 motif is specific to the Cic-S isoform and is present only in dipterans. To get further insight into the evolution of this element, we examined the structure of the *cic* locus in different insect taxa, focusing on the region that spans the alternatively spliced exons of *Drosophila cic-S* and *cic-L* transcripts. We were able to perform these analyses given the high conservation of peptide sequences encoded by these alternative exon junctions (Fig. 3B). We found a similar *cic* genomic organization in *Drosophila* and four distant species of lower dipterans: *Clogmia albipunctata*, *Culex pipiens*, *Anopheles gambiae* and *Aedes aegypti* (Fig. 6A; S2 Fig.), implying that this organization was already present in an early common ancestor of dipterans. In contrast, a different structure, which lacks the first *cic-S* exon, is apparent across representative species of non-dipteran taxa, including *Bombyx mori* (Lepidoptera), *Tribolium castaneum* (Coleoptera), *Apis mellifera* (Hymenoptera), and *Acyrthosiphon pisum* (Hemiptera) (Fig. 6A, B). In this configuration, the two exons encoding the N2-L motif of Cic-L proteins are frequently separated by short (<150 pb) introns, which do not contain the first *cic-S* exon encoding the N2 repressor motif (LYLQCLL) nor its upstream promoter region (Fig. 6A). Thus, while we cannot rule out the possibility that other short isoforms of Cic exist in non-dipteran species (e.g. expressed from other alternative promoters within *cic*), a form equivalent to dipteran Cic-S (containing the N2 motif) is clearly absent in those species. Therefore, the simplest interpretation of these genomic organizations is that the Cic-S isoform and its N2 motif originated after the expansion of the above *cic-L* intron during the early radiation of dipterans; an alternative scenario, where the Cic-S isoform was already present in early insects, appears much less likely, since this would involve the independent loss of this isoform in each of the non-dipteran branches examined.

These findings indicate that Cic-L represents the ancestral isoform of Cic in insects that gave rise to Cic-S in dipterans. To further test the significance of this evolutionary change, we compared the activities of the *Drosophila* Cic-L and Cic-S isoforms in early embryogenesis. The function of Cic-L has not been studied at the molecular level, and it is even unclear whether it functions as a repressor [2]. To assay Cic-L repressor activity in the early embryo, we generated a transgene expressing Cic-L under the control of the maternal *cic-S* promoter (Fig. 6C; Materials and Methods). This construct drives efficient accumulation of Cic-L in blastoderm nuclei (Fig. 6D), but does not rescue the embryonic *cic* phenotype (Fig. 6F, H), indicating that it cannot replace Cic-S in repressing the terminal gap genes. Since Cic-L lacks the N2 motif, we then tested a Cic-L derivative carrying the N2 sequence inserted N-terminal to the HMG-box (Fig. 6C). Strikingly, this protein ($Cic\text{-}L^{N2}$) showed significant, although not complete, rescue of the embryonic *cic* mutant phenotype (Fig. 6E, G, I). This indicates that the *Drosophila* Cic-S and Cic-L isoforms have very different molecular activities, and that evolution of the N2 motif represented a key innovation for Cic repressor function in the early embryo.

Discussion

We have shown that Cic proteins exhibit both Gro-dependent and -independent activities, and that this functional diversity is associated with the origin of the Cic-S isoform and the N2 motif in dipterans, approximately 250 million years ago. By comparison, other functional attributes of Cic such as their sensitivity to RTK signaling and their binding to specific sites in DNA, are more broadly conserved and therefore probably more ancient. For example, the MAPK-interacting domain of *Drosophila* Cic (C2) is clearly recognizable outside the dipterans [7], and Cic is downregulated by RTK signaling in mammalian cells [15,19]. Thus, while Cic proteins may have long served as sensors of RTK signaling, their mechanisms of repression appear to have evolved and adapted to fulfill new Cic functions in distinct transcriptional contexts. Below, we discuss the significance and implications of the newly evolved mechanism of Cic repression in fly embryogenesis.

Our results indicate that prior to the origin of dipterans, Cic was present in insects as a Cic-L-like isoform that lacked the N2 motif. Clearly, *Drosophila* Cic-L cannot function in the early embryo unless it carries the N2 motif from Cic-S (Fig. 6). This suggests that evolution of the N2 motif dramatically altered the mechanism of Cic repression by establishing a novel association with Gro. How, then, did the N2 motif appear? The comparison of different insect *cic* genes suggests that the N2 motif originated along with the Cic-S isoform, possibly through genomic rearrangements of intronic *cic-L* sequences that created a shorter *cic-S* transcript and subsequent evolution of a functional N2 motif via random drift. In this regard, it has been argued that short peptide sequences such as the WRPW and eh1 Gro-interacting motifs may be particularly easy to evolve by simple drift [29,30].

The N2 motif is different from the WRPW and eh1 motifs, and we still do not know its precise mechanism of action. By analogy to the WRPW and eh-1 motifs, which bind the central pore of the Gro β propeller, it is possible that N2 also recognizes this region of Gro. If this is correct, the N2 motif should adopt a conformation across the pore that is insensitive to the *MB41* mutation, just like another Gro mutation, *MB31*, prevents binding of WRPW but not eh-1 to the pore [28]. Another, non-exclusive possibility is that

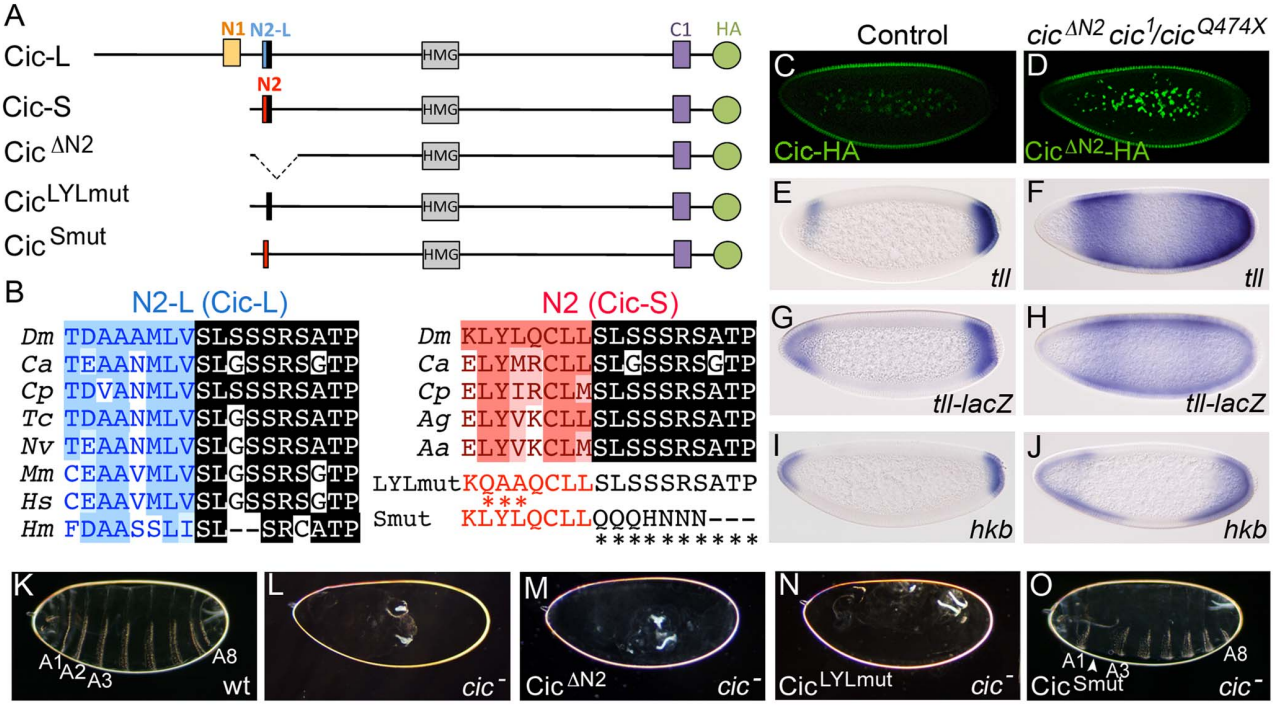


Fig. 3. The N2 motif is essential for Cic embryonic function. (A) Diagram of *Drosophila* Cic-L and Cic-S isoforms and three derivatives carrying mutations in the N2 motif. Cic-L and Cic-S are generated via use of alternative promoters and splicing sites, which produce different N-terminal domains. At the site of alternative splicing, Cic-L and Cic-S contain two different conserved motifs, N2-L and N2, which include different N-terminal sequences (shown in blue and red, respectively) and a common C-terminal peptide (highlighted in black). Other conserved domains (including the N1 and C1 domains of unknown function) are also indicated. The proteins are shown with an HA tag (green) to allow their visualization in transgenic embryos (see also below). (B) Alignment of Cic N2-L and N2 sequences from different species. *Dm*, *Drosophila melanogaster*; *Ca*, *Clogmia albipunctata*; *Cp*, *Culex pipiens*; *Tc*, *Tribolium castaneum*; *Nv*, *Nasonia vitripennis*; *Mm*, *Mus musculus*; *Hs*, *Homo sapiens*; *Hm*, *Hydra magnipapillata; Ag*, *Anopheles gambiae*; *Aa*, *Aedes aegypti*. Two different mutations of the N2 peptide (LYLmut and Smut) are also shown below the N2 alignment. (C and D) Expression of Cic and Cic$^{\Delta N2}$ proteins tagged with the HA epitope in embryos stained with anti-HA antibody; note that both proteins appear downregulated at the embryo poles. (E-M) mRNA expression patterns of *tll*, *tll-lacZ* and *hkb* in wt (E, G, I) and *cic* mutant (*cic*^{1}/*cic*^{Q474X}) embryos expressing Cic$^{\Delta N2}$ (F, H, J). Cuticle phenotypes of the same genetic backgrounds are shown in K and M, respectively; panel L shows a control *cic*^{1}/*cic*^{Q474X} mutant cuticle. (N and O) Cuticle phenotypes of *cic*^{1}/*cic*^{Q474X} mutant embryos expressing the Cic^{LYLmut} (N) and Cic^{Smut} (O) derivatives; only Cic^{Smut} rescues the *cic* phenotype, except for mild segmental defects (arrowhead). A1-A8, abdominal segments 1–8.
doi:10.1371/journal.pgen.1004902.g003

N2 binds the Gro β propeller with the help of auxiliary proteins. Consistent with this idea, the WRPY motif of Runx proteins binds very weakly to Gro and this interaction depends on other accessory proteins in vivo [28,31,32].

What could be the functional and evolutionary significance of the new Gro-dependent mechanism of Cic repression? It seems logical to assume that Cic employs qualitatively different mechanisms of repression in the embryo (via Gro) than when acting in other contexts (presumably with other corepressors; see below). We suggest that the combined activities of Cic-S and Gro may have facilitated the evolution of the complex transcriptional network regulated by Torso signaling in modern fly embryos. This network comprises multiple Cic target genes, including *tll* and *hkb*, whose boundaries of expression are regulated by Torso-dependent gradients of Cic repression at the embryo poles [9,10,33,34]. In this system, Gro itself appears to exert a regulatory function beyond its obvious role as a component of the repression machinery. Indeed, Gro is directly phosphorylated and functionally downregulated in response to Torso signaling [35,36], and even modest changes in Gro protein levels significantly affect the threshold concentrations at which Cic represses *tll* and *hkb* [37]. This suggests a model where Torso signaling controls the expression of Cic target genes via coordinate activity gradients of both Cic and Gro. These overlapping gradients might serve as a fail-safe mechanism to ensure the correct spatiotemporal response of target genes, buffering against random perturbations in either gradient. Furthermore, Gro is a highly versatile corepressor capable of functioning in different contexts of recruitment [22–24,38], which may explain the ability of Cic to regulate multiple targets simultaneously. For example, *tll* and *hkb* are activated by different mechanisms that are either dependent (*hkb*) or independent (*tll*) of Lilliputian, a component of the super elongation complex (SEC) [39,40], implying that Gro is capable of counteracting both activation mechanisms. Thus, the acquisition of Gro-mediated repression by Cic may have facilitated the precise, coordinated regulation of Cic target genes in response to Torso signaling.

In contrast, Gro is mostly dispensable for other Cic functions in the wing and the follicular epithelium (Fig. 1). The Cic-S isoform is sufficient for both of these functions [4,5,7], raising the possibility that Cic-S acts through other corepressors in those tissues. One potential candidate is the *Drosophila* ortholog of mammalian Atxn1 (dAtxn1; [41]). In mammals, Atxn1 and the related factor Ataxin1-Like (Atxn1L; also known as Brother of ATXN1, BOAT) potentiate Cic-S repressor activity in cultured cells [14,42], and directly interact with a short motif of Cic that is

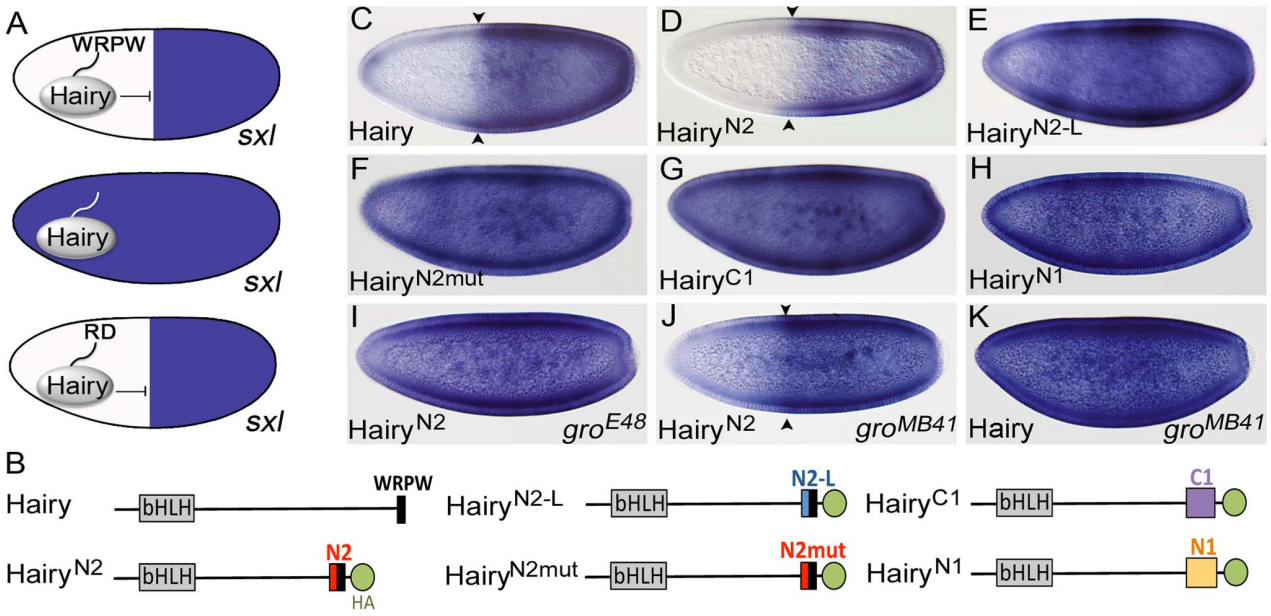


Fig. 4. N2 is a Gro-dependent repressor motif. (A) Schematic representation of the *Sxl* repression assay. In this assay, expression of Hairy under the control of the *hb* promoter in the anterior region of the embryo leads to repression of *Sxl* transcription (blue) in females (top). Repression depends on the WRPW motif of Hairy (middle), but replacement of this motif with autonomous repressor domains (RD) restores repression function (bottom). (B) Diagram of Hairy and Hairy fusion constructs tested in the *Sxl* assay; all fusions carry a C-terminal HA tag (see Materials and Methods). (C-K) Effects of Hairy constructs on *Sxl* expression in otherwise wild-type or *gro* mutant embryos; all images correspond to *Sxl* expression in female embryos. Arrowheads indicate borders of transcriptional repression. Note that Hairy^{N2} does not cause complete repression of *Sxl* in *gro*^{MB41} embryos (J); this may reflect a loss of function, in this genetic background, for endogenous Hairy-related factors such as Deadpan that normally contribute to *Sxl* repression [46].
doi:10.1371/journal.pgen.1004902.g004

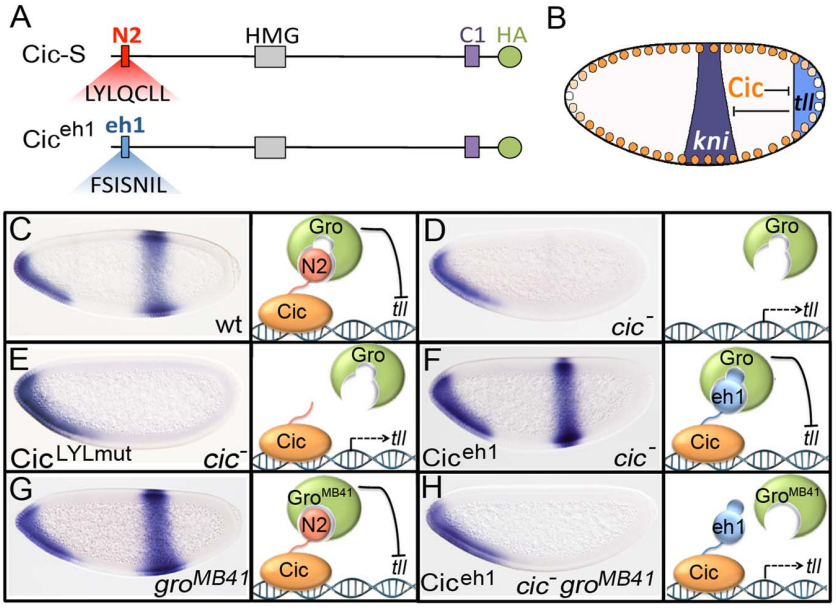


Fig. 5. The N2 motif of Cic recruits Gro to the terminal patterning system. (A) Diagram of Cic and Cic^{eh1} proteins; Cic^{eh1} carries the eh1 motif from *Drosophila* Engrailed instead of N2 and is tagged with an HA epitope at the C-terminus. (B) Schematic representation of cross-repressive interactions between Cic, *tll* and *kni* in the early blastoderm. (C-H) mRNA expression patterns of *kni* in wild-type (C), *cic* (D, E, F), *gro*^{MB41} (G) and *cic gro*^{MB41} (H) mutant backgrounds expressing the Cic^{LYLmut} (E) and Cic^{eh1} (F, H) products. A model diagram depicting the interactions of N2 and eh1 motifs with Gro proteins and the resulting repressor activities is shown next to each embryo; for simplicity, the interaction between N2 and Gro is modeled as being direct (see Discussion). The *cic* maternal mutant genotypes are *cic*^{1} for panels D, F and H, and *cic*^{1}/*cic*^{Q474X} for panel E.
doi:10.1371/journal.pgen.1004902.g005

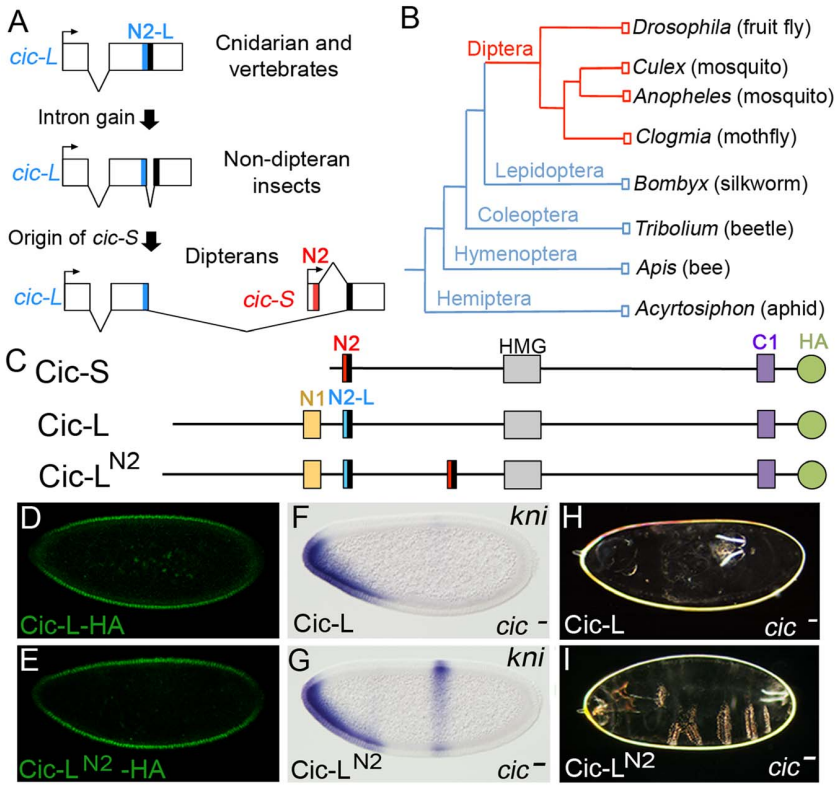


Fig. 6. Recent origin of the N2 motif in dipterans. (A) Schematic representation of proposed steps giving rise to the Cic-S isoform and the N2 motif. The diagrams show the regions spanning the alternatively spliced exons of *cic-S* and *cic-L* transcripts in selected species, which are represented graphically and are not drawn to scale. The size of introns splitting the N2-L coding sequences in non-dipteran species are as follows: *Bombyx mori* (758 bp), *Tribolium castaneum* (50 bp), *Apis mellifera* (86 bp) and *Acyrthosiphon pisum* (129 bp) (see also main text). The conserved protein motifs encoded at relevant exon junctions are shaded in red, blue or black as in Fig. 3. (B) Insect phylogeny illustrating the presence of the Cic-S and N2 motifs in dipterans (red). (C) Diagram of the Cic-S, Cic-L and Cic-L^{N2} proteins; Cic-L^{N2} carries the N2 motif inserted within a poorly conserved sequence of Cic. Cic-L and Cic-L^{N2} were expressed with an HA tag at the C-terminus. (D and E) Expression of HA-tagged Cic-L (D) and Cic-L^{N2} (E) proteins in embryos stained with anti-HA antibody. (F and G) mRNA expression patterns of *kni* in cic^{1}/cic^{Q474X} mutant embryos expressing Cic-L (F) and Cic-L^{N2} (G); only Cic-L^{N2} leads to significant rescue of the central *kni* stripe. (H and I) Cuticle phenotypes of the same genetic backgrounds shown in F and G, respectively.
doi:10.1371/journal.pgen.1004902.g006

conserved in *Drosophila* Cic-S [14,21]. dAtxn1 has been mainly studied in models of SCA1 pathogenesis [41]. Thus, future studies should examine whether dAtxn1 also mediates Cic repressor functions in development.

Finally, our results suggest that mammalian Cic proteins probably function independently of Gro, unless they have evolved other specific Gro-interacting motifs different from N2. Similarly, the mammalian Cic-S isoform must have originated independently of the dipteran Cic-S isoform, resulting in coincidental presence of Cic-S isoforms in both taxa. It will be interesting to determine whether mammalian Cic-S and Cic-L proteins also exhibit differential functional properties in their ability to regulate gene expression.

Materials and Methods

Drosophila genetics and transgenic lines

The following alleles were used: cic^{1} [3], cic^{Q474X} [8], gro^{E48}, gro^{MB36} and gro^{MB41} [28]. *cic* mutant embryos were obtained from cic^{1} or cic^{1}/cic^{Q474X} females, except in the experiments presented in Fig. 5D, F and H, which involved the generation of mosaic females whose germlines were homozygous for cic^{1} using the FRT/ovoD system [43]. All *gro* embryos were derived via the FRT/ovoD system. Transgenic lines were established by P-element-mediated transformation or using the ΦC31-based integration system [44]. The *hb-h* and $hb\text{-}h^{N2}$ transgenes cause high levels (>98%) of female lethality and were maintained in males, either using an attached X chromosome [C(1)M3] (for X-chromosome insertions) or unbalanced (for autosomal insertions). In contrast, the $hb\text{-}h^{N2\text{-}L}$, $hb\text{-}h^{N2mut}$, $hb\text{-}h^{C1}$ and $hb\text{-}h^{N1}$ transgenes do not cause female lethality, even when present in two copies.

DNA constructs

Cic-expressing transgenes were based on the original *cic* rescue construct [3], which contains the *cic-S* transcription unit flanked by its natural 5′ and 3′ regulatory sequences, and were assembled in *pCaSpeR4* or *pattB* vectors. The $Cic^{\Delta N2}$ construct lacks amino acids 3–77 of Cic-S. Cic^{ch1} contains the sequence VPLAFSISNIL instead of FQDFELGAKLYLQCLL. The Cic-L isoform used in this work is the product of cDNA *LD17181* (GenBank accession number BT100233), a fully sequenced clone identified by the Berkeley *Drosophila* Genome Project (see S2 Fig.). The LD17181 product (LD17181p) was expressed from the ATG initiator codon present in the *cic-S* rescue construct, by replacing the sequence encoding amino acids

4–19 of Cic-S with the sequence encoding amino acids 3–487 of LD17181p; note that amino acid 20 of Cic-S corresponds to amino acid 488 of LD17181p. Cic-L^{N2} was constructed by inserting an N2-containing fragment (residues 4-35 of Cic-S) at amino acid position 852 of LD17181p. All Cic derivatives have a triple HA tag (YPYDVPDYA) inserted in the same position, corresponding to amino acid 1398 of Cic-S. Hairy fusion proteins contain amino acids 1-268 of Hairy fused to the following Cic sequences: amino acids 3-35 ($Hairy^{N2}$), and 1308-1396 ($Hairy^{C1}$) of Cic-S, and amino acids 376-437 ($Hairy^{N1}$) and 468-503 ($Hairy^{N2\text{-}L}$) of LD17181p. $Hairy^{N2mut}$, contains the sequence AYAQCLASQ instead of LYLQCLLSL.

Embryo analyses

Embryos were fixed in 4% formaldehyde-PBS-heptane using standard procedures. In situ hybridizations were performed using digoxigenin-UTP labeled antisense RNA probes, and anti-digoxygenin antibodies conjugated to alkaline phosphatase (Roche). Immunodetection of HA-tagged Cic proteins was performed using monoclonal antibody 12CA5 (Roche) at 1:400 dilution and secondary Alexa488-conjugated antibodies (Molecular Probes). Cuticle preparations were mounted in 1:1 Hoyer's medium/lactic acid and cleared overnight at 60°C.

Supporting Information

S1 Fig Expression of Hairy chimeras inactive in the *Sxl* assay. (A-D) Expression of $Hairy^{N2\text{-}L}$, $Hairy^{N2mut}$, $Hairy^{C1}$ and $Hairy^{N1}$ proteins under the control of the *hb* promoter (see Fig. 4). All proteins are readily detected by anti-HA immunostaining, indicating that their inability to repress *Sxl* is not due to inefficient accumulation in the embryo.
(TIF)

S2 Fig Structure of the *Drosophila cic* locus and two main transcripts, *cic-S* and *cic-L*, expressed from alternative promoters. White and grey boxes indicate transcribed untranslated regions and coding sequences, respectively. Sequences encoding the N1, N2, N2-L, HMG-box and C1 domains are highlighted in color. The structure of the *cic-L* transcript corresponds to the *LD17181* cDNA (see Materials and Methods). The sequence of the first exon and its immediate upstream region is shown below to indicate the positions of the annotated transcription initiation site (TIS, bent arrow) and the 5′ end of *LD17181* (arrowhead). The position of the TIS is based on RNA-seq profiles generated by the modENCODE project [47]. The translated peptide sequence is also shown in bold, with residues encoded by *LD17181* highlighted in red; thus, the LD17181p product is 5 amino acid shorter than the corresponding predicted Cic-L protein (1871 vs. 1876 residues, respectively). Genomic sequences from *Drosophila erecta* (*De*) and *Drosophila yakuba* (*Dy*) are aligned below the *melanogaster* (*Dm*) sequence; note that both species contain in-frame stop codons (asterisks) immediately upstream of the N-terminal methionine, supporting the predicted initiation of translation.
(TIF)

Acknowledgments

We thank A. Olza for assistance with *Drosophila* injections, E. Jiménez-Guri and J. Jaeger for providing *cic* genomic sequences from *C. albipunctata*, S. Shvartsman, A. Veraksa, B. Jennings, A. Brehm, and D. Ish-Horowicz for discussions, M. Wainwright, B. Jennings and the Bloomington *Drosophila* Research Center for fly stocks, and S. Shvartsman for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: MF LA SA CN SGC ZP GJ. Performed the experiments: MF LA NS SA CN RG. Analyzed the data: MF LA NS SA CN RG SGC ZP GJ. Wrote the paper: SGC ZP GJ.

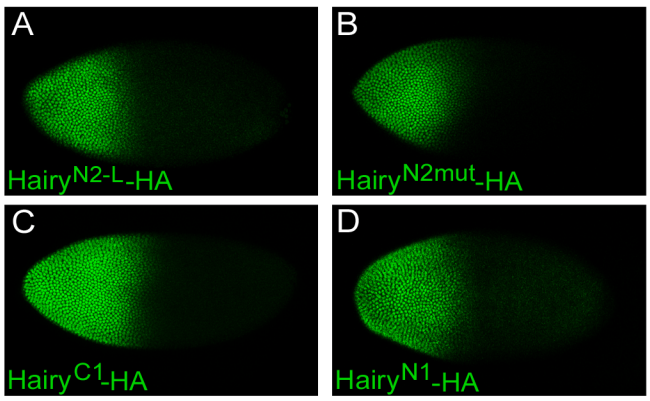
References

1. Lemmon MA, Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. Cell 141: 1117–1134.
2. Jiménez G, Shvartsman SY, Paroush Z (2012) The Capicua repressor - a general sensor of RTK signaling in development and disease. J Cell Sci 125: 1383–1391.
3. Jiménez G, Guichet A, Ephrussi A, Casanova J (2000) Relief of gene repression by Torso RTK signaling: role of *capicua* in *Drosophila* terminal and dorsoventral patterning. Genes Dev 14: 224–231.
4. Goff DJ, Nilson LA, Morisato D (2001) Establishment of dorsal-ventral polarity of the *Drosophila* egg requires capicua action in ovarian follicle cells. Development 128: 4553–4562.
5. Roch F, Jiménez G, Casanova J (2002) EGFR signalling inhibits Capicua-dependent repression during specification of *Drosophila* wing veins. Development 129: 993–1002.
6. Atkey MR, Lachance JF, Walczak M, Rebello T, Nilson LA (2006) Capicua regulates follicle cell fate in the *Drosophila* ovary through repression of *mirror*. Development 133: 2115–2123.
7. Astigarraga S, Grossman R, Diaz-Delfin J, Caelles C, Paroush Z, et al. (2007) A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. EMBO J 26: 668–677.
8. Tseng AS, Tapon N, Kanda H, Cigizoglu S, Edelmann L, et al. (2007) Capicua regulates cell proliferation downstream of the receptor tyrosine kinase/Ras signaling pathway. Curr Biol 17: 728–733.
9. Löhr U, Chung HR, Beller M, Jackle H (2009) Antagonistic action of Bicoid and the repressor Capicua determines the spatial limits of *Drosophila* head gene expression domains. Proc Natl Acad Sci U S A 106: 21695–21700.
10. Ajuria L, Nieva C, Winkler C, Kuo D, Samper N, et al. (2011) Capicua DNA-binding sites are general response elements for RTK signaling in *Drosophila*. Development 138: 915–924.
11. Andreu MJ, González-Pérez E, Ajuria L, Samper N, Gonzalez-Crespo S, et al. (2012) Mirror represses *pipe* expression in follicle cells to initiate dorsoventral axis formation in *Drosophila*. Development 139: 1110–1114.
12. Andreu MJ, Ajuria L, Samper N, González-Pérez E, Campuzano S, et al. (2012) EGFR-dependent downregulation of Capicua and the establishment of *Drosophila* dorsoventral polarity. Fly 6: 234–239.
13. Grimm O, Sanchez Zini V, Kim Y, Casanova J, Shvartsman SY, et al. (2012) Torso RTK controls Capicua degradation by changing its subcellular localization. Development 139: 3962–3968.
14. Lam YC, Bowman AB, Jafar-Nejad P, Lim J, Richman R, et al. (2006) ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. Cell 127: 1335–1347.
15. Fryer JD, Yu P, Kang H, Mandel-Brehm C, Carter AN, et al. (2011) Exercise and genetic rescue of SCA1 via the transcriptional repressor Capicua. Science 334: 690–693.
16. Bettegowda C, Agrawal N, Jiao Y, Sausen M, Wood LD, et al. (2011) Mutations in *CIC* and *FUBP1* contribute to human oligodendroglioma. Science 333: 1453–1455.
17. Lee Y, Fryer JD, Kang H, Crespo-Barreto J, Bowman AB, et al. (2011) ATXN1 protein family and CIC regulate extracellular matrix remodeling and lung alveolarization. Dev Cell 21: 746–757.
18. Kawamura-Saito M, Yamazaki Y, Kaneko K, Kawaguchi N, Kanda H, et al. (2006) Fusion between *CIC* and *DUX4* up-regulates *PEA3* family genes in Ewing-like sarcomas with t(4;19)(q35; q13) translocation. Hum Mol Genet 15: 2125–2137.
19. Dissanayake K, Toth R, Blakey J, Olsson O, Campbell DG, et al. (2011) ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription factors via the transcriptional repressor capicua. Biochem J 433: 515–525.
20. Lim J, Crespo-Barreto J, Jafar-Nejad P, Bowman AB, Richman R, et al. (2008) Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. Nature 452: 713–718.
21. Kim E, Lu H, Zoghbi HY, Song J (2013) Structural basis of protein complex formation and reconfiguration by polyglutamine disease protein Ataxin-1 and Capicua. Genes Dev 27: 590–595.
22. Jennings BH, Ish-Horowicz D (2008) The Groucho/TLE/Grg family of transcriptional co-repressors. Genome Biol 9: 205.
23. Turki-Judeh W, Courey AJ (2012) Groucho: a corepressor with instructive roles in development. Curr Top Dev Biol 98: 65–96.
24. Mannervik M (2014) Control of *Drosophila* embryo patterning by transcriptional co-regulators. Exp Cell Res 321: 47–57.

25. Paroush Z, Wainwright SM, Ish-Horowicz D (1997) Torso signalling regulates terminal patterning in *Drosophila* by antagonising Groucho-mediated repression. Development 124: 3827–3834.
26. Parkhurst SM, Bopp D, Ish-Horowicz D (1990) X:A ratio, the primary sex-determining signal in *Drosophila*, is transduced by helix-loop-helis proteins. Cell 63: 1179–1191.
27. Jiménez G, Paroush Z, Ish-Horowicz D (1997) Groucho acts as corepressor for a subset of negative regulators, including Hairy and Engrailed. Genes Dev 11: 3072–3082.
28. Jennings BH, Pickles LM, Pearl LH, Ish-Horowicz D (2006) Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. Mol Cell 22: 645–655.
29. Hittinger CT, Carroll SB (2008) Evolution of an insect-specific Groucho-interaction motif in the Engrailed selector protein. Evol Dev 10: 537–545.
30. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134: 25–36.
31. Aronson BD, Fisher AL, Blechman K, Caudy M, Gergen JP (1997) Groucho-dependent and -independent repression activities of Runt domain proteins. Mol Cell Biol 17: 5581–5587.
32. Canon J, Banerjee U (2003) In vivo analysis of a developmental circuit for direct transcriptional activation and repression in the same cell by a Runx protein. Genes Dev 17: 838–843.
33. Chen H, Xu Z, Mei C, Yu D, Small S (2012) A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes. Cell 149: 618–629.
34. Helman A, Lim B, Andreu MJ, Kim Y, Shestkin T, et al. (2012) RTK signaling modulates the Dorsal gradient. Development 139: 3032–3039.
35. Cinnamon E, Helman A, Ben-Haroush Schyr R, Orian A, Jiménez G, et al. (2008) Multiple RTK pathways downregulate Groucho-mediated repression in *Drosophila* embryogenesis. Development 135: 829–837.
36. Helman A, Cinnamon E, Mezuman S, Hayouka Z, Von Ohlen T, et al. (2011) Phosphorylation of Groucho mediates RTK feedback inhibition and prolonged pathway target gene expression. Curr Biol 21: 1102–1110.
37. Turki-Judeh W, Courey AJ (2012) The unconserved Groucho central region is essential for viability and modulates target gene specificity. PLoS One 7: e30610.
38. Cinnamon E, Paroush Z (2008) Context-dependent regulation of Groucho/TLE-mediated repression. Curr Opin Genet Dev 18: 435–440.
39. Tang AH, Neufeld TP, Rubin GM, Muller HA (2001) Transcriptional regulation of cytoskeletal functions and segmentation by a novel maternal pair-rule gene, *lilliputian*. Development 128: 801–813.
40. Luo Z, Lin C, Schilatifard A (2012) The super elongation complex (SEC) family in transcriptional control. Nat Rev Mol Cell Biol 13: 543–547.
41. Tsuda H, Jafar-Nejad P, Patel AJ, Sun Y, Chen HK, et al. (2005) The AXH domain of Ataxin-1 mediates neurodegeneration through its interaction with Gfi-1/Senseless proteins. Cell 122: 633–644.
42. Crespo-Barreto J, Fryer JD, Shaw CA, Orr HT, Zoghbi HY (2010) Partial loss of Ataxin-1 function contributes to transcriptional dysregulation in spinocerebellar ataxia type 1 pathogenesis. PLoS Genet 6: e1001021.
43. Chou TB, Noll E, Perrimon N (1993) Autosomal *P[ovoD1]* dominant female-sterile insertions in *Drosophila* and their use in generating germ-line chimeras. Development 119: 1359–1369.
44. Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific ΦC31 integrases. Proc Natl Acad Sci U S A 104: 3312–3317.
45. Housden BE, Terriente-Felix A, Bray SJ (2014) Context-dependent enhancer selection confers alternate models of notch regulation on *argos*. Mol Cell Biol 34: 664–672.
46. Younger-Shepherd S, Vaessin H, Bier E, Jan LY, Jan YN (1992) *deadpan*, an essential pan-neural gene encoding an HLH protein, acts as a denominator in *Drosophila* sex determination. Cell 70: 911–922.
47. Graveley BR, Brooks A, Carlson JW, Duff MO, Landolin JM, et al. (2011) The developmental transcriptome of *Drosophila melanogaster*. Nature 471: 473–479.

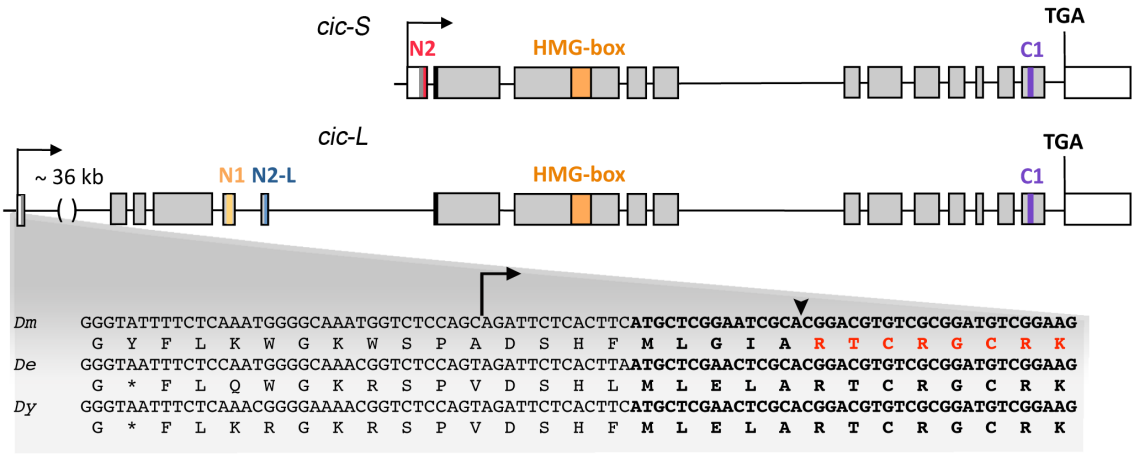
Supporting information

S1 Fig.



S1 Fig. Expression of Hairy chimeras inactive in the *Sxl* assay. (A-D) Expression of Hairy^{N2-L}, Hairy^{N2mut}, Hairy^{C1} and Hairy^{N1} proteins under the control of the *hb* promoter (see Fig. 4). All proteins are readily detected by anti-HA immunostaining, indicating that their inability to repress *Sxl* is not due to inefficient accumulation in the embryo.

S2 Fig.



S2 Fig. Structure of the *Drosophila cic* locus and two main transcripts, *cic-S* and *cic-L*, expressed from alternative promoters. White and grey boxes indicate transcribed untranslated regions and coding sequences, respectively. Sequences encoding the N1, N2, N2-L, HMG-box and C1 domains are highlighted in color. The structure of the *cic-L* transcript corresponds to the *LD17181* cDNA (see Materials and Methods). The sequence of the first exon and its immediate upstream region is shown below to indicate the positions of the annotated transcription initiation site (TIS, bent arrow) and the 5' end of *LD17181* (arrowhead). The position of the TIS is based on RNA-seq profiles generated by the modENCODE project [47]. The translated peptide sequence is also shown in bold, with residues encoded by *LD17181* highlighted in red; thus, the LD17181p product is 5 amino acid shorter than the corresponding predicted Cic-L protein (1871 vs. 1876 residues, respectively). Genomic sequences from *Drosophila erecta* (*De*) and *Drosophila yakuba* (*Dy*) are aligned below the *melanogaster* (*Dm*) sequence; note that both species contain in-frame stop codons (asterisks) immediately upstream of the N-terminal methionine, supporting the predicted initiation of translation.

PUBLICACIÓN 2: Using CRISPR-Cas9 to study ERK signaling in *Drosophila*

Autores:

Marta Forés, Aikaterini Papagianni, Laura Rodríguez-Muñoz, Gerardo Jiménez

Referencia:

In *ERK Signaling: Methods and Protocols* (ed. G. Jiménez), pp. 353-365. New York, NY: Springer New York (2017)

Resumen:

La ingeniería genómica usando la tecnología CRISPR-Cas9 está revolucionando la investigación biomédica. La técnica CRISPR-Cas9 permite la edición precisa de genes en una amplia variedad de células y organismos, acelerando los estudios moleculares a través de la mutagénesis dirigida, etiquetado de epítopos y otras modificaciones. En este artículo, ilustramos la metodología CRISPR-Cas9 centrándonos en Cic, un represor transcripcional directamente fosforilado e inactivado por ERK/MAPK. Específicamente, hemos usado CRISPR-Cas9 para delecionar un motivo de unión de ERK a la proteína Cic de *Drosophila*, generando un mutante insensible a ERK de este importante sensor de la señal.

Chapter 26

Using CRISPR-Cas9 to Study ERK Signaling in *Drosophila*

Marta Forés, Aikaterini Papagianni, Laura Rodríguez-Muñoz, and Gerardo Jiménez

Abstract

Genome engineering using the clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated nuclease 9 (Cas9) technology is revolutionizing biomedical research. CRISPR-Cas9 enables precise editing of genes in a wide variety of cells and organisms, thereby accelerating molecular studies via targeted mutagenesis, epitope tagging, and other custom genetic modifications. Here, we illustrate the CRISPR-Cas9 methodology by focusing on Capicua (Cic), a nuclear transcriptional repressor directly phosphorylated and inactivated by ERK/MAPK. Specifically, we use CRISPR-Cas9 for targeting an ERK docking site of *Drosophila* Cic, thus generating ERK-insensitive mutants of this important signaling sensor.

Key words CRISPR, Cas9, RTK signaling, ERK, MAPK, Capicua, Docking site, *Drosophila*

1 Introduction

With hundreds of genomes sequenced to date, and massively parallel sequencing of single individuals, inbred strains and transcriptomes from multiple tissues and biological conditions, the field of biomedicine is well ahead into the post-genomic era. The current challenge, therefore, is to manage and exploit this wealth of information to unravel the myriad regulatory functions encoded in the genome and ultimately understand the complex transitions from genotype to phenotype. It is in this context that genome-editing techniques relying on zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and, more recently, the type II clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated nuclease 9 (Cas9) system have revolutionized the entire field of biology. By enabling the introduction of custom changes in genomic sequences, these methods are facilitating highly diverse, sophisticated, and cost-effective applications across basic and applied disciplines.

Gerardo Jiménez (ed.), *ERK Signaling: Methods and Protocols*, Methods in Molecular Biology, vol. 1487,
DOI 10.1007/978-1-4939-6424-6_26,

The first basic step during genome engineering is the introduction of double-strand breaks (DSBs) at selected sites in the genome, which are then resolved by two main repair pathways [1–3] (Fig. 1). The primary pathway used by all eukaryotic cells, nonhomologous end joining (NHEJ), is an error-prone mechanism where the break

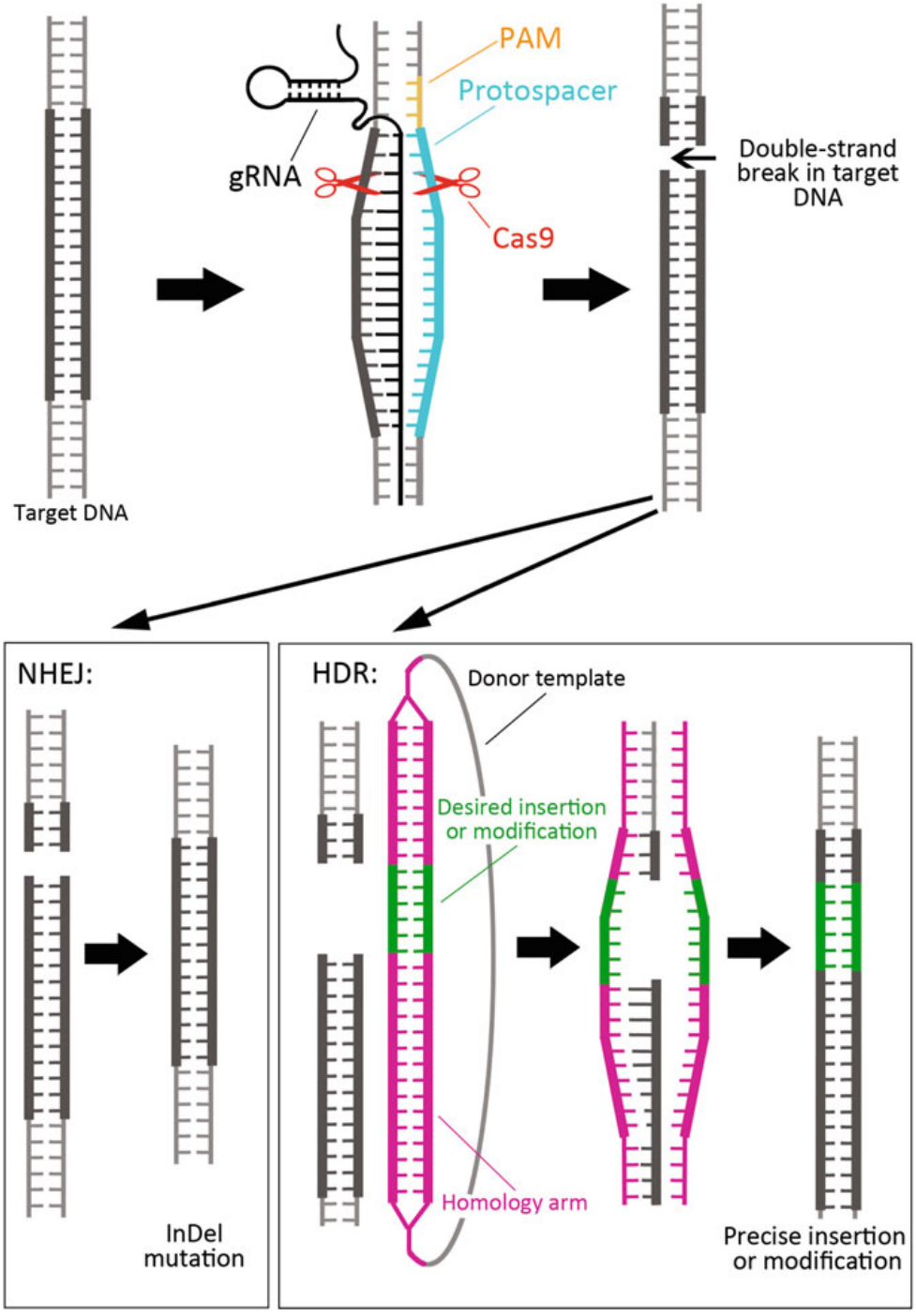


Fig. 1 Overview of the CRISPR/Cas system and two major DNA damage repair pathways. The gRNA contains a guide sequence portion that recognizes the target DNA through RNA–DNA base pairing. Following the production of a DSB usually 3 bp upstream of the PAM, the broken DNA is repaired by NHEJ or HDR

ends are directly ligated and this often causes small mutagenic insertions and/or deletions (InDels). The second pathway, homology-directed repair (HDR), restores the break by precisely copying a template sequence that bears homology across the DSB site. The donor template can be present naturally in the cell (e.g. the sister chromatid), or can be supplied experimentally to exploit the HDR system and incorporate custom modifications at the break site.

The major improvement offered by the CRISPR-Cas9 system lies in the ability to induce localized DSBs in a highly efficient manner (Fig. 1). CRISPR-Cas9 is based on a bacterial immune system where DNA sequences derived from bacteriophages and other exogenous genetic elements become integrated in clustered repeats (CRISPR loci) in the bacterial genome, are then transcribed along with CRISPR-associated (*cas*) nuclease-encoding and other non-coding genes, and the resulting RNAs and nucleases form a complex that cleaves foreign DNAs complementary to the CRISPR RNAs [4–7]. This system has been engineered to function with just two components, a single synthetic guide RNA (gRNA) containing ~20 nucleotides (nt) of complementarity to a genomic target site of interest and the Cas9 enzyme from *Streptococcus pyogenes* [6], and demonstrated to work in eukaryotic cells [8–12] (Fig. 1). The choice of the target sequence (called protospacer) is rather flexible, requiring only a 3′ protospacer-adjacent motif (PAM) recognized by Cas9 [13]; the PAM for *S. pyogenes* Cas9 is 5′-NGG-3′, a motif found on average every 8–12 base pairs (bp) in the human genome. To date, CRISPR has been adopted to modify the genome of numerous animal and plant species [14, 15] and has enabled extraordinary applications such as the genome-wide inactivation of retroviral copies in the pig genome [16], the spreading of autocatalytic mutations in flies [17], and the generation of wheat plants resistant to powdery mildew disease [18].

In this chapter, we illustrate the use of CRISPR-Cas9 to study Ras-ERK/MAPK signaling in *Drosophila* [19–25]. Specifically, we focus on Capicua (Cic), a conserved Ras-ERK signaling sensor with crucial roles in development and human diseases [26]. Cic is a Sox-type HMG-box protein that represses genes induced by ERK signaling, maintaining them silenced in the absence of signaling. When ERK is activated, it directly phosphorylates and downregulates Cic and this relieves repression of its target genes. This control can be disrupted by mutating a conserved ERK docking site in Cic, which creates a dominant repressor form that escapes ERK-mediated downregulation [27, 28]. However, the effects of this mutation have been studied only using transgenes expressing one isoform of Cic (a short isoform named Cic-S), not in the context of the native *cic* locus, which also encodes a long isoform (Cic-L) [26]. We describe the isolation of such endogenous mutations via CRISPR-Cas9.

2 Materials

2.1 Drosophila Strains

1. ΦC31 system host strain, e.g. *y*[1] *v*[1] *P{nos-phiC31}X; P{CaryP} attP40* (currently available from the Bloomington *Drosophila* Stock Center at Indiana University with number 25709).
2. *vermilion* mutant strains for recovery and mapping of ΦC31 positive transformants, e.g. *v*[1] (currently available from the Bloomington Stock Center with number 137) and *v*[1]*; sna*[Sco]*/SM6a* (available from our lab upon request).
3. Cas9-expressing line, e.g. *y*[2] *cho*[2] *v*[1]*; attP40{nos-Cas9}/CyO* (currently available from the National Institute of Genetics at Shizuoka, Japan, with name CAS-0001) (http://www.shigen.nig.ac.jp/fly/nigfly/cas9/index.jsp) [21].

2.2 Reagents and Solutions

1. gRNA expression vector, e.g. *pCFD3-dU6:3gRNA* [25] (available from Addgene with number 49410).
2. Custom complementary oligonucleotides for inserting the 20-bp guide sequence into the gRNA vector.
3. BbsI restriction enzyme and T4 DNA ligase.
4. 20× SSC: 3 M NaCl, 0.3 M tri-sodium citrate, pH 7. Dissolve 175.3 g of NaCl and 88.3 g of tri-sodium citrate dihydrate in 800 mL of distilled water. Adjust the pH to 7 with NaOH or HCl and make up to 1 L with water.
5. Squishing buffer: 10 mM Tris–HCl pH 8.2, 1 mM EDTA, 25 mM NaCl. Before use, add Proteinase K to a final concentration of 200 μg/mL (from a 20 mg/mL stock stored at –20 °C).

3 Methods

The ERK docking motif of Cic (designated C2) is located in the region shared by the Cic-S and Cic-L isoforms (Fig. 2a; *see* **Note 1**). Based on previous genetic analyses of *cic* [27–32], we anticipated two possible types of mutations resulting from CRISPR-mediated targeting of the C2 coding sequence. Frameshift mutations resulting in truncated Cic-S and Cic-L products should behave as recessive loss-of-function alleles. In contrast, in-frame mutations causing substitutions or deletions at the C2 sequence should behave as dominant (or semidominant) gain-of-function alleles. Since Cic acts downstream of the EGFR signaling pathway which is essential for zygotic viability, we also considered the possibility that inactivation of the C2 motif (caused by in-frame mutations) might be dominant lethal and therefore impossible to analyze in vivo. With all this in mind, we based our approach on the generation of InDels via NHEJ, since this would allow the recovery of different C2 alleles, with frameshift mutations serving as an internal control.

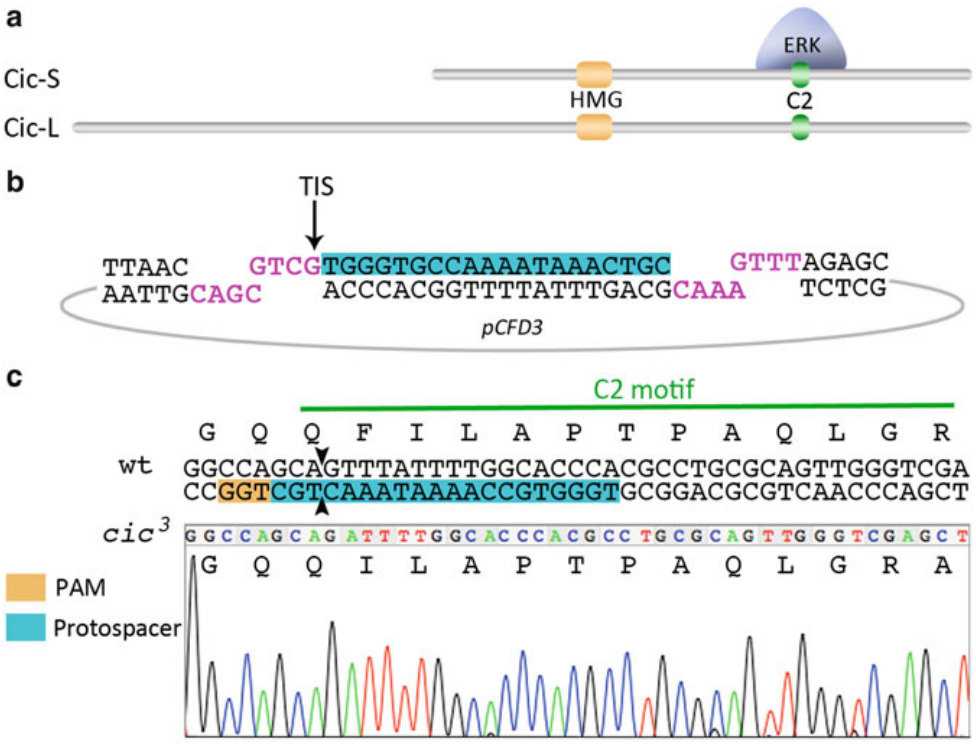


Fig. 2 Isolation of a CRISPR-induced mutation in the C2 motif of Cic. (**a**) Diagram of the Cic-S and Cic-L isoforms showing the HMG-box and C2 ERK docking site. Cic-L contains an N-terminal extension of unknown molecular function [26, 43, 44]. (**b**) Schematics of the gRNA expression construct used for targeting the C2 coding sequence. Shown is the BbsI linker with the guide sequence highlighted in *blue*. TIS: transcription initiation site of the gRNA. (**c**) Coding sequence of the C2 motif indicating the protospacer and PAM. The predicted cleavage site of Cas9 is indicated by *arrowheads*. A sequencing chromatogram of a PCR product amplified from a *cic*^{3} homozygous fly is shown below; note the loss of a single phenylalanine (F) codon

3.1 Generation of gRNA-Expressing Transgenic Lines

To design a gRNA targeting the C2 coding sequence, we first selected a protospacer followed by a PAM motif (NGG) within that sequence (Fig. 2b). We then prepared the corresponding gRNA expression construct using the *pCFD3-dU6:3gRNA* vector [25], which contains an 80-bp gRNA core sequence flanked by promoter and 3′ genomic sequences from the *Drosophila U6:3* spliceosomal snRNA gene. The plasmid contains a BbsI restriction-site cassette between the promoter and gRNA core, which can be replaced by a guide sequence of choice. In addition, the plasmid carries a bacterial attachment sequence (*attB*) for ΦC31 integrase-mediated recombination with phage attachment sites (*attP*) inserted in a host fly strain [33], and a *vermilion* marker gene for selection of transformants.

The steps to generate transgenic gRNA-expressing flies are as follows (*see* **Note 2**):

1. Select an 18- to 20-bp protospacer sequence (adjacent to a PAM motif) and order complementary oligonucleotides that will leave cohesive BbsI ends (Fig. 2b, *see* **Note 3**).

2. Anneal the oligonucleotides in a microfuge tube at a final concentration of 0.5 μg/μL for each oligonucleotide and 1× SSC in a total volume of 20 μL. Place the tube in a 1-L beaker of water at 80–85 °C and allow to cool to room temperature (*see* **Note 4**).
3. Ligate the annealed oligonucleotides to BbsI-linearized *pCFD3-dU6:3gRNA* vector, using 2–3 different concentrations of annealed oligos in the range of 0.5–2 ng/μL (~1:1000 dilution relative to the initial concentration of 1 μg/μL).
4. Transform *E. coli* cells with the ligation mixture (*see* **Note 5**).
5. Identify positive transformants by colony PCR using one of the above oligonucleotides and a flanking primer annealing ~200 bp upstream or downstream of the cloning site.
6. Grow a positive colony, isolate the plasmid DNA, and verify the construct by sequencing across the cloning junctions.
7. For fly germ-line transformation, prepare high-quality *pCFD3-gRNA* plasmid DNA and inject it into a suitable host strain for ΦC31-mediated site-specific integration, such as y^1 v^1 *P{nos-phiC31}X; P{CaryP}attP40* [33] (*see* **Note 6**).
8. Recover hatched G0 flies and cross them individually to a *vermilion* (*v*) strain (e.g. v^1) for identification of transformants in the F1 progeny (*see* **Note 7**).
9. Confirm the site-specific chromosomal integration of the transgene by crossing the positive transformants (preferably males) with appropriate balancer stocks in a v^1 background—e.g. v^1; sna^{Sco}/*SM6a* for mapping of *attP40* second chromosome insertions. Also, by crossing males and females carrying the insertion over the balancer chromosome, it should be possible to recover homozygous transgenic flies and use them to establish a homozygous stock.

3.2 Isolation of Mutants via NHEJ

Once a homozygous gRNA line has been established, it can be crossed to a Cas9-expressing line to obtain founder animals that carry both transgenes and can transmit NHEJ-mediated mutations to their progeny.

1. Cross 3–5 homozygous gRNA flies with 3–5 Cas9-expressing (for example, y^2 cho^2 v^1; *attP40{nos-Cas9}/CyO*) flies of the opposite sex (*see* **Note 8**).
2. Collect founder males expressing the *U6-gRNA* and *nos-Cas9* transgenes and cross them individually to virgin females carrying a balancer for the chromosome on which the targeted locus is located (chromosome 3 in the case of *cic*) (*see* **Note 9**).
3. From the above cross, select heterozygous flies (preferably males) potentially carrying an NHEJ-induced mutation over the balancer chromosome and cross them individually to a balancer stock to save the potential mutation (*see* **Note 10**).

4. Two–three days later, when the first larval progeny begin to hatch, take the parent fly carrying a potential mutation and mash it in 40 μL of Squishing buffer (*see* **Note 11**). Incubate at 37 °C for 30 min.
5. Inactivate the Proteinase K by heating the sample to 95 °C for 3 min. Then use 1 μL of the sample for a PCR reaction with a primer combination that amplifies across the targeted genomic sequence (*see* **Notes 12** and **13**).
6. Sequence the PCR product and keep the progeny that carries an interesting mutation (from **step 4**) (*see* **Notes 14** and **15**).

3.3 Analysis of cic Mutations in the C2-Coding Region

Using the above protocol, we have obtained a set of mutations affecting the C2-coding sequence. Out of over 50 alleles recovered, approximately 81% were frameshift mutations that behave as *cic* loss-of-function mutations. These mutations consisted in short InDels (mainly deletions) ranging from 1 to 20 bp in length, with up to 45% of the mutations removing more than 10 bp. In contrast, the remaining 19% in-frame InDel mutations were generally shorter, usually affecting 1–2 codons flanking the Cas9 cut site. For example, one mutation (hereafter designated cic^3) caused deletion of a single conserved F residue (QFI) in the C2 motif (Fig. 2c). Another mutation removed a longer sequence (DADGQQF), but this sequence only partly overlaps the highly conserved C2 core (QFILAPTPAQLG).

Our interpretation of this mutational pattern is that short deletions and substitutions within C2 do not completely abolish the binding of C2 to ERK, permitting partial ERK-dependent downregulation of Cic that is sufficient for zygotic viability. In contrast, loss of >2 amino acid residues within the C2 core presumably causes dominant lethality and the corresponding mutations are simply not recovered. This is somewhat unexpected given that transgenic flies expressing a Cic-S construct lacking the C2 motif are perfectly viable—although this construct causes female sterility by interfering with maternal Cic-S function in the early embryo (see below) [27]. It thus appears that complete loss of C2 function in the context of both Cic-S and Cic-L isoforms is completely deleterious, suggesting that Cic-L contributes to some essential function(s) that is normally antagonized by RTK signaling during development. The molecular mechanisms of this postulated function and its associated dominant lethality caused by C2 deletions are currently unknown.

Our preliminary analyses indicate that the cic^3 allele confers excess Cic function resulting from lower sensitivity to ERK-induced downregulation. To illustrate this idea, we present three examples in connection with different RTK signaling processes (*see* **Note 16**). First, the eggs laid by cic^3 females show clear defects in the respiratory appendages that differentiate on the dorsal-anterior surface of the chorion (Fig. 3a). These structures are specified by a dorsal

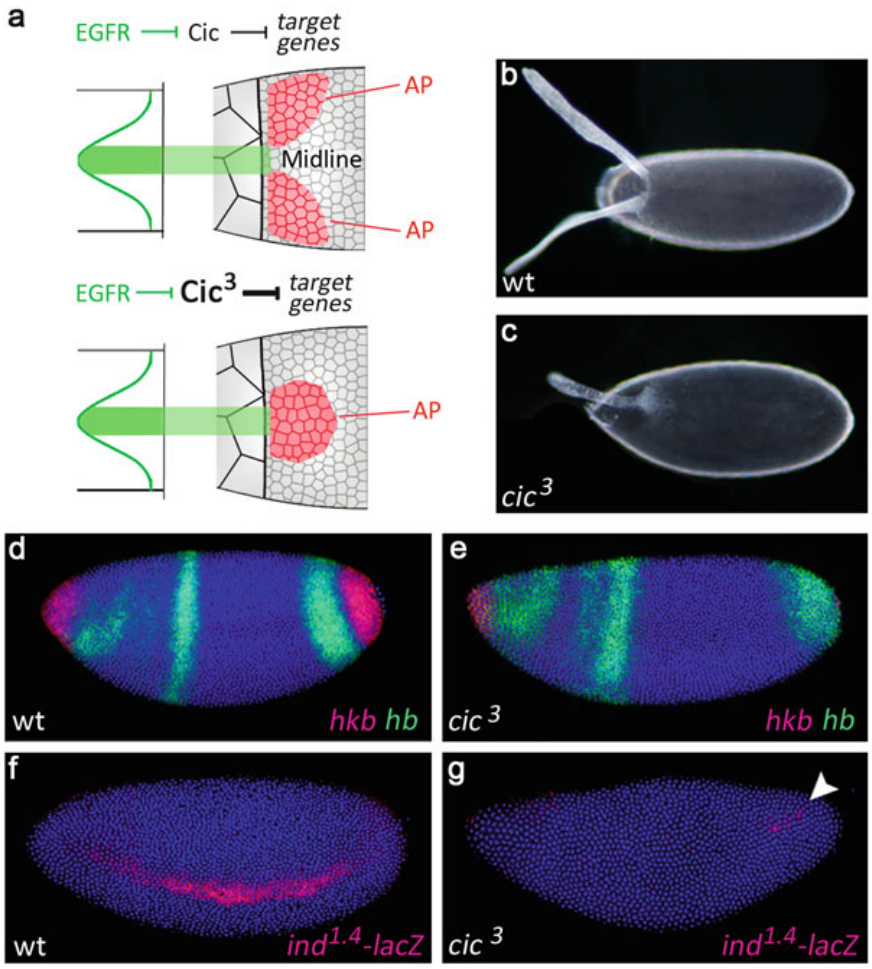


Fig. 3 Phenotypes caused by the *cic*^{3} allele. (**a**) Simplified representation of the patterning mechanisms controlling dorsal appendage differentiation. The *cartoons* show dorsal views of the follicular epithelium, where two dorsal appendage (DA) primordia (AP) form on either side of the midline; the large cells on the left are nurse cells. These structures are controlled by a gradient of EGFR signaling, such that high levels of activity (*green area*) induce the midline fate, whereas lower levels specify the flanking AP. Thus, increasing EGFR signaling produces widely spaced DAs [45], whereas reducing it causes a loss of midline fate and fused DAs [37]. Although how EGFR signaling controls target gene expression in follicle cell nuclei is not fully understood, a direct mechanism involves downregulation of Cic repressor activity through ERK phosphorylation [27]. This, in turn, relieves repression of Cic target genes in follicle cells [27, 28, 36]. Since the Cic^{3} protein should be partly resistant to this downregulation (indicated by *bold-type* face), it behaves like a reduction in EGFR signaling, possibly causing inappropriate repression of target genes [27, 28]. One likely target of Cic in this context is *pointed*, which is essential for establishing the midline [46–48]. The system is subject to additional mechanisms of control, including feedback loops and inputs from other pathways, which are not considered here. (**b**, **c**) Eggshells laid by wild-type and *cic*^{3} homozygous females; note the fusion of DAs in the latter. (**d**, **e**) Patterns of *hkb* and *hb* mRNA expression in wild-type (**d**) and maternally mutant *cic*^{3} (**e**) embryos; note the absent (*hkb*) and shifted (*hb*) domains of expression in the *cic*^{3} mutant. (**f**, **g**) Expression of *ind* in wild-type (**f**) and maternally mutant *cic*^{3} (**g**) embryos, visualized using a transgenic *ind*$^{1.4}$*-lacZ* reporter under the control of *ind* regulatory sequences [40, 49]. Note the strongly reduced expression of *ind*$^{1.4}$*-lacZ* in the mutant, with only minimal staining along the stripe (*arrowhead*)

gradient of EGFR signaling in the follicular epithelium of the female ovary [34, 35]. Peak levels of EGFR signaling in dorsal positions specify a midline fate devoid of appendage material, whereas lower levels in dorsolateral regions induce a respiratory appendage on either side of the midline. This control is mediated, at least in part, through EGFR-dependent downregulation of Cic in dorsal follicle cells [27, 28, 36] (Fig. 3a). Thus, mutations disrupting EGFR signaling lead to a fusion or complete loss of chorion appendages [37], whereas reduced Cic function causes the opposite effect, i.e. lateral broadening of the appendages and differentiation of appendage-like material in ventral positions [30, 36]. As shown in Fig. 3b, c, cic^3 homozygous females lay eggs that exhibit partial or complete fusion of appendages, a phenotype consistent with decreased sensitivity of the mutant Cic^3 protein to EGFR-mediated downregulation.

Second, the embryos derived from cic^3 homozygous females show clear defects in anteroposterior patterning. These phenotypes are related to the role of Cic downstream of the maternal Torso RTK pathway, which is active at the embryonic poles and specifies the anterior and posterior terminal body structures [26]. During its activation in the early embryo, Torso signaling induces zygotic genes such as *huckebein* (*hkb*), which is locally expressed at the poles (Fig. 3d). This induction depends on ERK-mediated downregulation of Cic, which acts as a repressor of *hkb* and maintains it silenced outside of the poles (where Torso is OFF) [29, 38]. In cic^3 maternally mutant embryos, the Cic^3 protein escapes this ERK-dependent control, leading to partial or complete repression of *hkb* at the poles and, consequently, embryonic lethality. This is illustrated in Fig. 3e, where *hkb* appears clearly repressed and there is a corresponding shift of *hunchback* (*hb*) expression towards the pole, since *hb* is normally repressed by the Hkb product (compare with Fig. 3d) (*see* **Note 17**).

Finally, we examined the phenotype of cic^3 maternal embryos in the context of embryonic dorsoventral patterning. This process depends in part on activation of the EGFR RTK pathway along a ventrolateral longitudinal stripe on each side of the embryo (Fig. 3f). This localized activity induces expression of *intermediate neuroblasts defective* (*ind*), a gene required for patterning the future nerve cord. As in the case of Torso signaling, this transcriptional induction depends on ERK-dependent relief of Cic repression [38–40]. Consistent with this model and the molecular nature of Cic^3, embryos expressing this mutant protein show greatly reduced expression of an *ind* reporter transgene (Fig. 3g), implying that binding of ERK to C2 is crucial for linking EGFR signaling to transcriptional activation of *ind*.

In summary, our CRISPR-induced mutations affecting the C2 ERK docking motif of Cic indicate an essential requirement of this motif during *Drosophila* development. We propose that C2 controls the activity of both Cic-S and Cic-L isoforms in multiple developmental contexts, thus acting as a major molecular sensor of Ras-ERK signaling in this organism.

4 Notes

1. The *cic* gene has a complex organization and probably encodes multiple Cic-S and cic-L isoforms (further details can be found at the FlyBase website). For simplicity, we only consider a single representative of each isoform type.
2. gRNA constructs can be provided as integrated transgenes (which are then crossed to Cas9-expressing lines) or by direct injection into embryos expressing Cas9. To facilitate the generation of multiple C2 mutations, we generated a transgenic gRNA line which could be easily crossed to Cas9 strains to produce as many founder animals as necessary.
3. Protospacers in the range of 18–20 bp have been shown to function effectively [41], though the actual cleavage efficiency is difficult to predict a priori. Ideally, the protospacer sequence should begin with a 5′ G, since transcription from the *U6:3* promoter begins with a G which is always present in the cohesive 5'-end of the BbsI cassette (this requirement applies to most gRNA vectors). This will ensure that the 18- to 20-nt guide sequence of the gRNA will be identical to the genomic protospacer sequence (Fig. 1). However, gRNAs starting with a 5′ G not present in the target sequence are also usually functional. In addition, it is advisable to evaluate any potential off-target effects of the selected guide sequences. In *Drosophila*, CRISPR-Cas9 appears to function with high specificity, but running similarity searches and off-target prediction analyses [42] will help reduce the risk of generating undesired mutations at other genomic positions. Finally, it is important to verify that the gRNA and Cas9 strains used in the experiments do not contain polymorphisms across the genomic target sequence of interest, since they would prevent efficient recognition by the gRNA.
4. We use 0.2- or 0.5-mL tubes which are placed inside a 50-mL Falcon tube mostly submerged in the water (using a foil or glass lid on the beaker).
5. It is advisable to set up a control ligation and transformation without insert. This transformation should produce very few colonies, since the two BbsI sites present in *pCFD3-dU6:3gRNA* produce incompatible ends that cannot self-ligate to recircularize the vector.
6. Plasmids intended for injection should be prepared using a midiprep or maxiprep kit (e.g. from Qiagen). The y^1 v^1 *P{nos-phiC31}X; P{CaryP}attP40* strain carries a *nos-PhiC31* transgene (expressing the ΦC31 integrase in germ cells under the control of 5′ and 3′ regulatory elements of the *nanos* gene) on the X chromosome, and the *attP40* landing site at position 25C6 on the second chromosome.

7. We typically establish 50–80 individual crosses using a single G0 fly and 3 v^1 flies of the opposite sex. This should produce at least one positive transformant, which will show a wild-type eye color and pigmented ocelli—as opposed to the bright red eye color and unpigmented ocelli characteristic of v^1 flies.
8. We usually cross gRNA-expressing males to *cas9* females, but the opposite cross should work as well (*see* for example [21]).
9. In general, we cross 8–12 founder males for analysis of their progeny.
10. The number of flies to be analyzed will depend on the efficiency of the CRISPR mutagenesis, and the nature of the desired mutations. For example, it should be easier to isolate a simple frameshift mutation in a coding sequence than a relative long in-frame deletion of five or more residues. We typically establish consecutive sets of crosses (e.g. 10–15 crosses each time) that we analyze until we identify the desired mutation(s).
11. Mash the fly for 5–10 s with a yellow tip containing 40 μL of Squishing buffer without expelling any liquid (sufficient liquid escapes from the tip). Then add the remaining solution to the squashed fly.
12. We tend to use this sample within 1–2 days of preparation.
13. In general, we amplify short fragments of 200–400 bp using 30 cycles.
14. Since the parent fly was heterozygous, the presence of a mutation will produce double peaks in the chromatogram. By identifying the wild-type sequence in those peaks, it is usually possible to infer the sequence of the mutant allele.
15. Alternatively, since PCR products from heterozygous mutant flies form DNA heteroduplexes at the final cycles of the PCR reaction, it is also possible to analyze these products using T7 Endonuclease I (T7EI), which will cleave the heteroduplex at the mismatch site and produce two bands in an agarose gel [21]. For this, 10 μL of the PCR product are treated with 0.5 μL of T7EI enzyme in a total volume of 20 μL containing 1× NEBuffer 2 (New England Biolabs) for 15 min at 37 °C. The reactions are then terminated by incubating on ice for 3 min and the samples are loaded in a gel along with 10 μL of untreated PCR product.
16. These phenotypes show a dose-dependent relationship with the number of mutant copies, being stronger in homozygous condition.
17. The pattern of *hb* expression in cic^3 embryos also suggests a certain level of Cic downregulation in those embryos, since complete loss of Torso signaling activity abolishes posterior *hb* expression. This probably reflects residual expression of *tailless*—another Cic target downstream of Torso—in cic^3 embryos, given that Tailless (indirectly) activates *hb* (*see* refs. 26, 27 for further details).

Acknowledgments

We thank A. Olza for *Drosophila* injections, N. Samper for experimental support, and F. Port, S. González-Crespo, Z. Paroush, M. Ruiz-Gómez, and A. Veraksa for discussions. This work was funded by grants from the Spanish Ministry of Science and Innovation (BFU2014-52863-P) and Fundació La Marató de TV3 (20131730). G.J. is an ICREA investigator.

References

1. Moynahan ME, Jasin M (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. Nat Rev Mol Cell Biol 11:196–207
2. Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem 79:181–211
3. Jackson SP, Bartek J (2009) The DNA-damage response in human biology and disease. Nature 461:1071–1078
4. Mojica FJM, Díez-Villaseñor C, García-Martínez J et al (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol 60:174–182
5. Deltcheva E, Chylinski K, Sharma CM et al (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471:602–607
6. Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821
7. Gasiunas G, Barrangou R, Horvath P et al (2012) Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A 109:E2579–E2586
8. Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823
9. Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. Science 339:823–826
10. Jinek M, East A, Cheng A et al (2013) RNA-programmed genome editing in human cells. eLife 2:e00471
11. Cho SW, Kim S, Kim JM et al (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31:230–232
12. Hwang WY, Fu Y, Reyon D et al (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31:227–229
13. Anders C, Niewoehner O, Duerst A et al (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513:569–573
14. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157: 1262–1278
15. Wright AV, Nuñez JK, Doudna JA (2016) Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. Cell 164:29–44
16. Yang L, Güell M, Niu D et al (2015) Genome-wide inactivation of porcine endogenous retroviruses (PERVs). Science 350:1101–1104
17. Gantz VM, Bier E (2015) The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. Science 348:442–444
18. Wang Y, Cheng X, Shan Q et al (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat Biotechnol 32:947–951
19. Gratz SJ, Cummings AM, Nguyen JN et al (2013) Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. Genetics 194:1029–1035
20. Bassett AR, Tibbit C, Ponting CP et al (2013) Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep 4:220–228
21. Kondo S, Ueda R (2013) Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. Genetics 195:715–721
22. Yu Z, Ren M, Wang Z et al (2013) Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. Genetics 195: 289–291

23. Ren X, Sun J, Housden BE et al (2013) Optimized gene editing technology for *Drosophila melanogaster* using germ line-specific Cas9. Proc Natl Acad Sci U S A 110:19012–19017
24. Gratz SJ, Ukken FP, Rubinstein CD et al (2014) Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. Genetics 196:961–971
25. Port F, Chen HM, Lee T et al (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. Proc Natl Acad Sci U S A 111:E2967–E2976
26. Jiménez G, Shvartsman SY, Paroush Z (2012) The Capicua repressor – a general sensor of RTK signaling in development and disease. J Cell Sci 125:1383–1391
27. Astigarraga S, Grossman R, Díaz-Delfín J et al (2007) A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. EMBO J 26:668–677
28. Andreu MJ, Ajuria L, Samper N et al (2012) EGFR-dependent downregulation of Capicua and the establishment of *Drosophila* dorsoventral polarity. Fly 6:234–239
29. Jiménez G, Guichet A, Ephrussi A et al (2000) Relief of gene repression by Torso RTK signaling: role of *capicua* in *Drosophila* terminal and dorsoventral patterning. Genes Dev 14:224–231
30. Goff DJ, Nilson LA, Morisato D (2001) Establishment of dorsal-ventral polarity of the *Drosophila* egg requires *capicua* action in ovarian follicle cells. Development 128:4553–4562
31. Roch F, Jiménez G, Casanova J (2002) EGFR signalling inhibits Capicua-dependent repression during specification of *Drosophila* wing veins. Development 129:993–1002
32. Tseng ASK, Tapon N, Kanda H et al (2007) Capicua regulates cell proliferation downstream of the receptor tyrosine kinase/Ras signaling pathway. Curr Biol 17:728–733
33. Bischof J, Maeda RK, Hediger M et al (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific φC31 integrases. Proc Natl Acad Sci U S A 104:3312–3317
34. Berg CA (2005) The *Drosophila* shell game: patterning genes and morphological change. Trends Genet 21:346–355
35. Cheung LS, Schüpbach T, Shvartsman SY (2011) Pattern formation by receptor tyrosine kinases: analysis of the Gurken gradient in *Drosophila* oogenesis. Curr Opin Genet Dev 21:719–725
36. Atkey MR, Boisclair Lachance JF, Walczak M et al (2006) Capicua regulates follicle cell fate in the *Drosophila* ovary through repression of *mirror*. Development 133:2115–2123
37. Schüpbach T (1987) Germ line and soma cooperate during oogenesis to establish the dorsoventral pattern of egg shell and embryo in *Drosophila melanogaster*. Cell 49:699–707
38. Ajuria L, Nieva C, Winkler C et al (2011) Capicua DNA-binding sites are general response elements for RTK signaling in *Drosophila*. Development 138:915–924
39. Garcia M, Stathopoulos A (2011) Lateral gene expression in *Drosophila* early embryos is supported by Grainyhead-mediated activation and tiers of dorsally-localized repression. PLoS One 6:e29172
40. Lim B, Samper N, Lu H et al (2013) Kinetics of gene derepression by ERK signaling. Proc Natl Acad Sci U S A 110:10330–10335
41. Ren X, Yang Z, Xu J et al (2014) Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in *Drosophila*. Cell Rep 9:1151–1162
42. Marx V (2014) Gene editing: how to stay on-target with CRISPR. Nat Methods 11:1021–1026
43. Lam YC, Bowman AB, Jafar-Nejad P et al (2006) ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. Cell 127:1335–1347
44. Forés M, Ajuria L, Samper N et al (2015) Origins of context-dependent gene repression by Capicua. PLoS Genet 11:e1004902
45. Neuman-Silberberg FS, Schupbach T (1994) Dorsoventral axis formation in *Drosophila* depends on the correct dosage of the gene *gurken*. Development 120:2457–2463
46. Boisclair Lachance JF, Fregoso Lomas M, Eleiche A et al (2009) Graded Egfr activity patterns the *Drosophila* eggshell independently of autocrine feedback. Development 136: 2893–2902
47. Zartman JJ, Kanodia JS, Cheung LS et al (2009) Feedback control of the EGFR signaling gradient: superposition of domain-splitting events in *Drosophila* oogenesis. Development 136:2903–2911
48. Jin Y, Ha N, Forés M et al (2015) EGFR/Ras signaling controls *Drosophila* intestinal stem cell proliferation via Capicua-regulated genes. PLoS Genet 11:e1005634
49. Stathopoulos A, Levine M (2005) Localized repressors delineate the neurogenic ectoderm in the early Drosophila embryo. Dev Biol 280:482–493

PUBLICACIÓN 3: Minibrain and Wings apart control organ growth and tissue patterning through down-regulation of Capicua

Autores:

Liu Yang, Sayantanee Paul, Kenneth G. Trieu, Lucas G. Dent, Francesca Froldi, **Marta Forés**, Kaitlyn Webster, Kellee R. Siegfried, Shu Kondo, Kieran Harvey, Louise Cheng, Gerardo Jiménez, Stanislav Y. Shvartsman, Alexey Veraksa

Referencia:

Proc Natl Acad Sci USA **113**, 10583-10588 (2016)

Resumen:

El represor transcripcional Capicua (Cic) controla el patrón tisular y restringe el crecimiento de órganos y, recientemente, se ha relacionado con varios tipos de cáncer. Cic ha emergido como un sensor primario de la señalización por debajo de la vía RTK/MAPK, pero cómo la actividad de Cic está regulada en diferentes contextos tisulares es algo que no se conoce del todo bien. Hemos visto que la quinasa Minibrain (Mnb) (ortólogo de la proteína de mamíferos DYRK1A), actúa a través de la proteína adaptadora Wings apart (Wap), interaccionando físicamente y fosforilando la proteína Cic. Mnb y Wap inhiben la función de Cic limitando su actividad como represor transcripcional. La regulación negativa de Cic a través de Mnb/Wap es necesaria para promover el crecimiento de múltiples órganos, incluyendo las alas, los ojos y el cerebro, y para un apropiado patrón de venación del ala. De este modo, hemos descubierto un mecanismo hasta ahora desconocido de inactivación de Cic por Mnb y Wap, los cuales operan independientemente del control de Cic a través de MAPK. De este modo, Cic funciona como un integrador de señales que son esenciales para el patrón tisular y el crecimiento de órganos. Finalmente, como DYRK1A y Cic exhiben actividades pro oncogénicas y supresoras tumorales respectivamente en oligodendroglioma, nuestros resultados abren la posibilidad que DYRK1A pueda también regular a la baja a CIC en células humanas.

Minibrain and Wings apart control organ growth and tissue patterning through down-regulation of Capicua

Liu Yang[a], Sayantanee Paul[a], Kenneth G. Trieu[a], Lucas G. Dent[b], Francesca Froldi[b], Marta Forés[c], Kaitlyn Webster[a], Kellee R. Siegfried[a], Shu Kondo[d], Kieran Harvey[b], Louise Cheng[b], Gerardo Jiménez[c,e], Stanislav Y. Shvartsman[f], and Alexey Veraksa[a,1]

[a]Department of Biology, University of Massachusetts, Boston, MA 02125; [b]Cell Growth and Proliferation Laboratory, Peter MacCallum Cancer Centre, East Melbourne, VIC 3002, Australia; [c]Instituto de Biología Molecular de Barcelona, Consejo Superior de Investigaciones Científicas, 08028 Barcelona, Spain; [d]Laboratory of Invertebrate Genetics, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka, Japan; [e]Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain; and [f]Department of Chemical and Biological Engineering and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544

Edited by Norbert Perrimon, Harvard Medical School/Howard Hughes Medical Institute, Boston, MA, and approved July 26, 2016 (received for review June 13, 2016)

The transcriptional repressor Capicua (Cic) controls tissue patterning and restricts organ growth, and has been recently implicated in several cancers. Cic has emerged as a primary sensor of signaling downstream of the receptor tyrosine kinase (RTK)/extracellular signal-regulated kinase (ERK) pathway, but how Cic activity is regulated in different cellular contexts remains poorly understood. We found that the kinase Minibrain (Mnb, ortholog of mammalian DYRK1A), acting through the adaptor protein Wings apart (Wap), physically interacts with and phosphorylates the Cic protein. Mnb and Wap inhibit Cic function by limiting its transcriptional repressor activity. Down-regulation of Cic by Mnb/Wap is necessary for promoting the growth of multiple organs, including the wings, eyes, and the brain, and for proper tissue patterning in the wing. We have thus uncovered a previously unknown mechanism of down-regulation of Cic activity by Mnb and Wap, which operates independently from the ERK-mediated control of Cic. Therefore, Cic functions as an integrator of upstream signals that are essential for tissue patterning and organ growth. Finally, because DYRK1A and CIC exhibit, respectively, prooncogenic vs. tumor suppressor activities in human oligodendroglioma, our results raise the possibility that DYRK1A may also down-regulate CIC in human cells.

minibrain | capicua | organ growth | DYRK1A | tissue patterning

The high mobility group-box transcriptional repressor protein Capicua (Cic) has been identified as a key regulator of tissue patterning and organ growth in multiple developmental contexts (1, 2). In *Drosophila*, Cic controls anteroposterior and dorsoventral embryonic polarity, the subdivision of the lateral ectoderm, and pattern formation in several tissues (1, 3–6). In addition, Cic negatively regulates the growth of imaginal discs and the midgut (7, 8). In humans, a single Cic ortholog (CIC) has been implicated in the neurodegenerative disease spinocerebellar ataxia 1 (SCA1) (9), and recently mutations in CIC have been found in the majority of oligodendroglioma cases, suggesting that CIC is a tumor suppressor (10–12).

In both *Drosophila* and mammals, Cic functions as a primary sensor of signaling downstream of the receptor tyrosine kinase (RTK)/extracellular signal-regulated kinase (ERK) pathway (2, 5–8, 13–16). According to the current model, activation of RTK signaling results in the accumulation of doubly phosphorylated activated ERK, which directly binds to and phosphorylates Cic (5). ERK-mediated Cic phosphorylation leads to a rapid relief of repression of Cic target genes, followed by a slower export from the nucleus and eventual cytoplasmic degradation (13, 17). The molecular details of these processes are unknown, although apparently each of them contributes to the overall down-regulation of Cic activity. Cic is also involved in a mutual regulatory relationship with the Hippo pathway, although regulation of Cic in this context appears to take place at the RNA level (18).

Here, we present the identification of the kinase Minibrain (Mnb) (19, 20) and an adaptor protein, Wings Apart (Wap) (20, 21), as Cic regulators that cooperate to phosphorylate Cic and restrict its repressor activity. We show that Mnb/Wap and ERK target different regions of the Cic protein for phosphorylation, and that inhibition of Cic activity by Mnb and Wap is required for the growth of several organs and for correct patterning of the wing. Our data suggest that Mnb/Wap-dependent down-regulation of Cic occurs in parallel to the RTK/ERK and Hippo signaling pathways. We propose that Cic functions as an integrator of upstream developmental signals, which together tightly control its activity. This mechanism is necessary for the proper execution of tissue patterning and regulation of organ growth.

Results

Wap and Mnb Interact with Cic. To identify Cic regulators, we used affinity purification/mass spectrometry (AP-MS) (22) to study the Cic protein interactome in *Drosophila* S2 cells and embryos. Embryonic *cic-Venus* was expressed at endogenous levels as part of a genomic rescue construct (13). We successfully recovered most of the known interactors of Cic, such as the *Drosophila* ERK ortholog Rolled (5), Ataxin-1 (9), and 14-3-3 proteins (16) (Fig. 1*A*). One of the prominent Cic interactors identified with high confidence in both cultured cells and embryos was Wap (also known as Riquiqui) (Fig. 1*A* and Dataset S1) (20, 21). Wap binds to the kinase Mnb (19, 20), and this interaction is

Significance

The transcriptional repressor protein Capicua (Cic) is a conserved regulator of organ growth and tissue patterning, and mutations in the CIC gene in humans result in the brain cancer oligodendroglioma. Cic activity is controlled by the receptor tyrosine kinase (RTK) signaling pathway. Here, we identify the kinase Minibrain (Mnb) and its adaptor Wings apart (Wap) as Cic regulators. Mnb and Wap bind to and phosphorylate the Cic protein, and inhibit the ability of Cic to repress gene expression. Mnb-dependent down-regulation of Cic is necessary for the proper growth of multiple organs and correct the patterning of tissues. Our results uncover a previously unknown mechanism of Cic regulation that acts in parallel to other growth-controlling pathways.

Author contributions: L.Y., K.H., G.J., S.Y.S., and A.V. designed research; L.Y., S.P., K.G.T., L.G.D., F.F., M.F., and K.W. performed research; S.K., K.H., L.C., G.J., and S.Y.S. contributed new reagents/analytic tools; L.Y., S.P., K.G.T., L.G.D., F.F., M.F., K.R.S., K.H., L.C., G.J., S.Y.S., and A.V. analyzed data; and L.Y. and A.V. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: alexey.veraksa@umb.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1609417113/-/DCSupplemental.

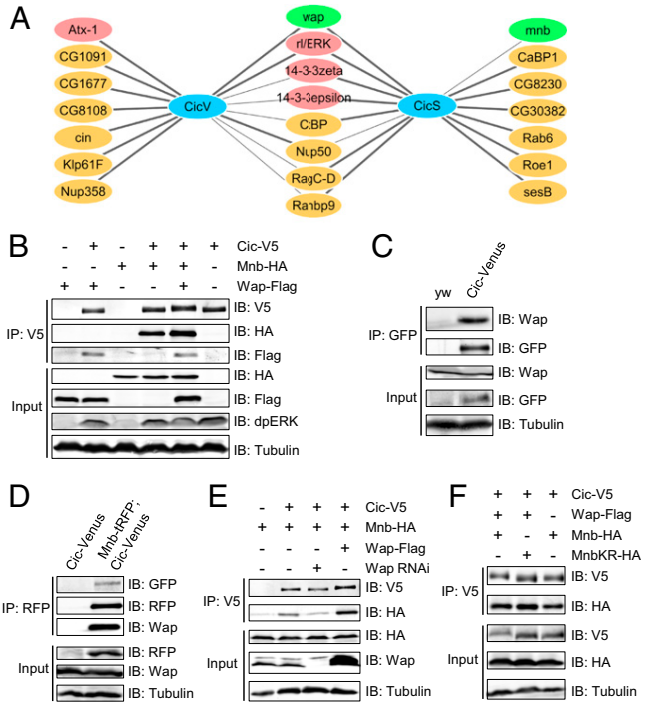


Fig. 1. Mnb and Wap physically interact with Cic, and Wap promotes the binding of Mnb to Cic. (*A*) The Cic protein interactome identified in *Drosophila* S2 cells (CicS) and embryos (CicV). Thick lines, highly significant interactions. A complete dataset is in Dataset S1. (*B*) Western blots showing coimmunoprecipitation of Cic, Mnb, and Wap in S2 cells. Endogenous dpERK is stabilized by Cic expression. (*C* and *D*) Coimmunoprecipitation of Cic, Mnb, and Wap in vivo using embryo lysates from *yw* (control), *cic-Venus*, or *cic-Venus* crossed with *mnb-tRFP*. (*E*) Wap is required and sufficient to bridge Cic and Mnb. (*F*) Cic mobility was changed when Cic was coexpressed with Wap and wild-type Mnb, but not kinase-dead Mnb (MnbKR).

conserved in mammals, because the Wap ortholog DDB1 and CUL4 associated factor 7 (DCAF7) forms a stable complex with the dual-specificity tyrosine phosphorylation-regulated kinase 1A (DYRK1A), which is an ortholog of Mnb (23). We have shown that Wap functions together with Mnb to regulate wing and leg tissue growth through the Hippo pathway (20). Our AP-MS experiments also identified four peptides of Mnb in the Cic-streptavidin binding peptide (SBP) pulldown in S2 cells (Fig. 1*A* and Dataset S1), suggesting that Wap, Mnb, and Cic form a protein complex. Coimmunoprecipitation in S2 cells using overexpressed proteins confirmed that Cic binds to both Wap and Mnb (Fig. 1*B*). To study the interactions between proteins expressed at endogenous levels in vivo, we generated tagged *mnb-tagRFP-T* (*mnb-tRFP*) and *wap-Venus* alleles by CRISPR/Cas9-mediated homologous recombination. Successful targeting was confirmed by RNAi (Fig. S1), and we found that both Mnb–tRFP and Wap–Venus were expressed throughout imaginal discs and the larval brain (Fig. S2). Endogenous Wap was detected in the Cic–Venus complex by using an anti-DCAF7 antibody (Fig. 1*C*), and both Wap and Cic–Venus were present in Mnb–tRFP complexes isolated from embryos in which Mnb–tRFP and Cic–Venus were coexpressed (Fig. 1*D*). Next, we asked whether Wap could serve as a bridge for the interaction between Mnb and Cic. RNAi depletion of *wap* in S2 cells led to a reduction in the binding of Cic to Mnb, whereas overexpression of Wap promoted the interaction (Fig. 1*E*). Collectively, these data suggest that Wap, Mnb, and Cic form a protein complex, with Wap likely serving as a bridging adaptor between Cic and Mnb.

Mnb Phosphorylates the Amino-Terminal Third of Cic in S2 Cells. Given that Mnb/DYRK1A is a kinase (19, 20), we asked whether Mnb could phosphorylate Cic. Cic mobility on an SDS/PAGE was reduced when Cic, Mnb, and Wap were coexpressed (Fig. 1 *B*, *E*, and *F*). Notably, the levels of activated ERK (dpERK) did not increase in this condition (Fig. 1*B*). In contrast, a kinase-dead mutant of Mnb (MnbKR) (20) failed to reduce Cic mobility (Fig. 1*F*), suggesting that the kinase activity of Mnb is required to reduce Cic mobility and that this modification is likely to be phosphorylation.

To determine which region of Cic was phosphorylated by Mnb, three Cic fragments (Cic1–3; Fig. 2*A*) were coexpressed with Wap in the presence or absence of Mnb in S2 cells. Only the amino-terminal fragment of Cic (Cic1, representing amino acids 1–453) was found to interact with Mnb (Fig. 2*B*). In addition, Mnb decreased the electrophoretic mobility of Cic1, but not Cic2 or Cic3 (Fig. 2*B*). Phos-tag gel analysis confirmed phosphorylation of Cic1 by wild-type but not kinase-dead Mnb (Fig. 2*C*). Next, we asked which residue(s) in Cic1 are phosphorylated by Mnb. Threonine 28 is part of a motif in Cic1 (RSATP) that closely matches the DYRK1A phosphorylation consensus RP(X)(S/T)P (24). Surprisingly, mutation of this residue (T28A) did not alter the phosphorylation pattern of Cic1 (Fig. 2*C*). To identify Cic residues that are phosphorylated by Mnb, Cic1 and Wap were coexpressed in S2 cells either with Mnb or MnbKR, Cic1 was purified, and its phosphorylation was analyzed by mass spectrometry. Four Cic1 residues (S41, S49, T89, and S91) were more highly phosphorylated by Mnb compared with MnbKR, with T89 and S91 phosphorylations found exclusively in the wild-type Mnb sample (Fig. 2*A* and Fig. S3). The S41 and S49 residues were also found to

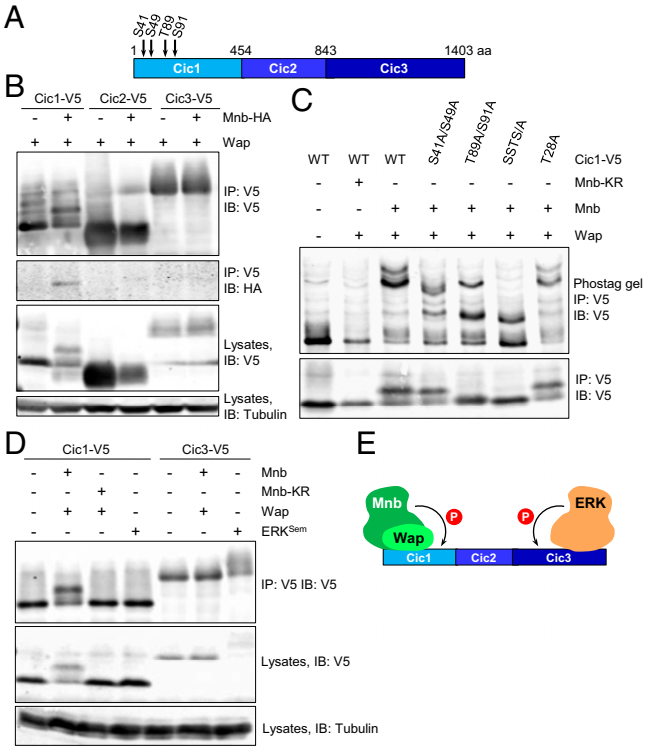


Fig. 2. Mnb and ERK target different regions of Cic for phosphorylation. (*A*) Schematic diagram of the three Cic fragments (Cic1, Cic2, and Cic3) with locations of phosphorylation sites. (*B*) Mnb interacts with and phosphorylates only the amino-terminal Cic fragment, Cic1. (*C*) Phos-tag gel analysis of Cic1 phosphorylation. (*C Bottom*) Regular SDS/PAGE. (*D*) Mnb phosphorylates region Cic1, whereas activated ERK (ERK^{Sem}) phosphorylates region Cic3. (*E*) Summary of Cic binding and phosphorylation data.

be phosphorylated in an unbiased global phosphoproteomic study in *Drosophila* embryos (25). Alanine substitutions of the four residues resulted in a reduction of phosphorylation of Cic1, with the most pronounced effect observed for a quadruple mutant, Cic-SSTS/A (Fig. 2*C*).

Our previous studies showed that region Cic3 includes an ERK docking site and is subject to ERK-mediated phosphorylation (5, 14, 26). To compare the activities of Mnb and ERK, we coexpressed Cic1 or Cic3 with Mnb or a constitutively active *Drosophila* ERK, ERK^{Sem} (27). Mnb could only reduce the electrophoretic mobility of Cic1 but not Cic3, whereas ERK^{Sem} only reduced the electrophoretic mobility of Cic3 but not Cic1 (Fig. 2*D*). Collectively, these results suggest that Mnb and ERK target different regions of Cic for phosphorylation: Wap facilitates Mnb-dependent phosphorylation of the amino-terminal third of Cic, whereas ERK targets the carboxyl-terminal region (Fig. 2*E*).

Mnb and Wap Reduce Cic Repressor Activity. Previous studies have shown that phosphorylation of Cic by ERK can result in down-regulation of Cic by lowering its repressor activity, protein level, or nuclear localization (5, 6, 17). We hypothesized that Mnb may exert similar effects. First, we used the CoinFLP-GAL4 system (28) to generate RNAi-depletion clones in the eye imaginal discs (Fig. S4*A*). As expected, we observed reduced levels of Cic protein in *UAS-cic-RNAi* clones (Fig. S4*B*). However, no obvious increase in Cic protein level or change in subcellular localization was found in CoinFLP-generated *UAS-mnb-RNAi* clones (Fig. S4*C*). Therefore, Mnb is unlikely to control Cic at the level of protein turnover or nuclear access.

We have shown that the relief of Cic repressor function by ERK does not necessarily require reduction in Cic protein levels (17). To assess whether Mnb could similarly affect Cic repressor activity, we used a reporter, *CUASC-lacZ*, which contains five GAL4 binding sites flanked on either side by two Cic binding motifs (Fig. 3*A*) (6). This reporter is only responsive to GAL4 in areas where Cic activity is inhibited, e.g., by RTK signaling (Fig. 3*B*). Uniform induction of GAL4 expression in the wing pouch under the control of the *C5-GAL4* driver (29) resulted in a localized activation of LacZ expression in prospective veins (Fig. 3*C*). This pattern results from epidermal growth factor receptor (EGFR)/ERK-mediated inactivation of Cic in these regions (Fig. 3*B*) (6). RNAi depletion of *cic* or overexpression of ERK^{Sem} throughout the wing pouch led to a much broader expression of LacZ (Fig. 3*D* and Fig. S5*B*), confirming that the normal restriction of the expression pattern of *CUASC-lacZ* to prospective veins is Cic-dependent. In contrast, overexpression of Cic resulted in the loss of LacZ expression in the vein L5 region (Fig. 3*E*, open arrowhead). Overexpression of Mnb or Wap induced a broader LacZ expression in the wing pouch (Fig. 3*F* and Fig. S5*C*). Conversely, RNAi depletion of *mnb* or *wap* under the control of *C5-GAL4* resulted in reduced LacZ expression, particularly in vein L5 (Fig. 3*G* and Fig. S5*D*). These data suggest that Mnb and Wap limit Cic repressor function in the wing disc. This contribution likely complements the regulation by ERK, which appears to be insufficient on its own, at least for vein L5 (Fig. 3*B*).

Mnb and Wap have been shown to phosphorylate and inhibit Warts, which results in elevated Yki activity (20). To test whether *CUASC-lacZ* expression was affected by Hippo signaling, we depleted the levels of the Yki-interacting transcription factor Scalloped (Sd), which is required for the activation of Yki targets (30). We observed that knockdown of *sd* using RNAi had no obvious effect on the expression of *CUASC-lacZ* (Fig. S5*E*), suggesting that Hippo signaling is not involved in the regulation of Cic repressor activity in this context. To further assess whether Mnb and Wap engage RTK/ERK signaling to control Cic, we analyzed dpERK levels in wing pouches expressing *mnb-RNAi* or *wap-RNAi*. We found that RNAi depletion of *wap* or *mnb* did not alter the dpERK pattern in wing discs (Fig. S5 *F–H*). This result is in agreement with the observation that overexpression of Mnb did not increase dpERK levels in S2 cells (Fig. 1*B*). We conclude that Mnb and Wap down-regulates Cic repressor activity independently from the RTK/ERK and the Hippo pathways.

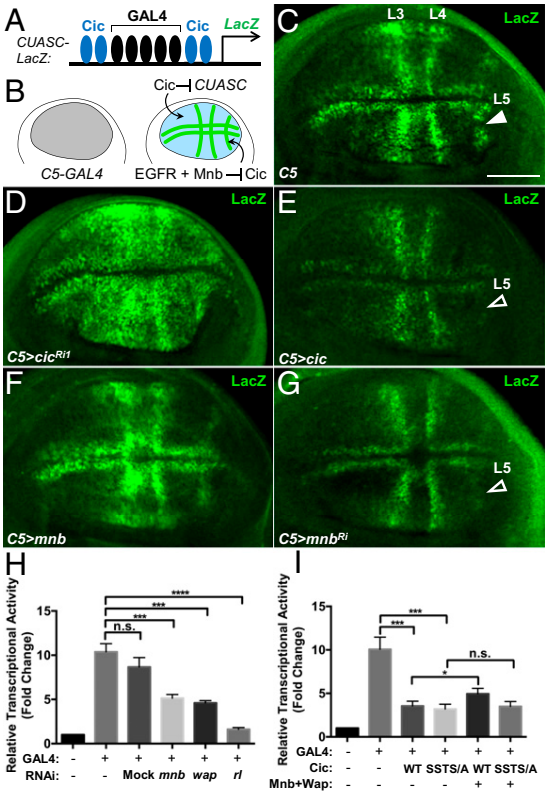


Fig. 3. Mnb reduces Cic repressor activity. (*A*) Diagram of the *CUASC-lacZ* reporter. (*B*) Summary diagram of expression patterns. (*C–G*) LacZ expression pattern resulting from *C5-GAL4*–directed activation of *CUASC-lacZ* in wing discs from control (*C*), *UAS-cic*^{RNAi1} (*D*), *UAS-cic* (*E*), *UAS-mnb* (*F*), and *UAS-mnb*^{RNAi} (*G*) larvae. (Scale bar: 50 μm.) (*H* and *I*) Luciferase assays using *CUASC-Luc* reporter in S2 cells. (*H*) *mnb*, *wap*, and *rl* (ERK) are required to limit the activity of Cic. (*I*) Mnb and Wap reduce transcriptional repressor activity of wild-type Cic, but not of the phosphorylation site mutant, Cic-SSTS/A. n.s., not significant, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, statistical significance was analyzed by using unpaired Student's *t* test. Error bars represent SD.

To directly address how Mnb and Wap affect Cic function as a transcriptional repressor, we studied the activity of a reporter, *CUASC-Luc*, which is controlled by GAL4 and Cic, in S2 cells (Fig. S5*I*). Transfection of GAL4 activated this reporter ~10-fold, and this activation was repressed by coexpression of Cic in a dose-dependent manner (Fig. S5*J*). Depletion of endogenous *mnb*, *wap*, or *rl* (ERK) by RNAi resulted in a reduction of reporter activity (Fig. 3H), suggesting that Mnb, Wap and ERK are required to limit the activity of Cic. We next tested whether Mnb and Wap could reduce the capacity of Cic to repress *CUASC-Luc* expression, and found that cotransfection of Cic with Mnb and Wap partially relieved Cic-mediated repression of this reporter (Fig. 3*I*). Whereas the Cic-SSTS/A mutant repressed the reporter gene expression to a similar level as wild-type Cic, coexpression of this mutant with Mnb and Wap did not affect its ability to repress *CUASC-Luc* (Fig. 3*I*). Collectively, these results indicate that Mnb and Wap reduce the activity of Cic as a transcriptional repressor, likely via Mnb-mediated phosphorylation of residues located in the amino terminus of the Cic protein.

DEVELOPMENTAL BIOLOGY

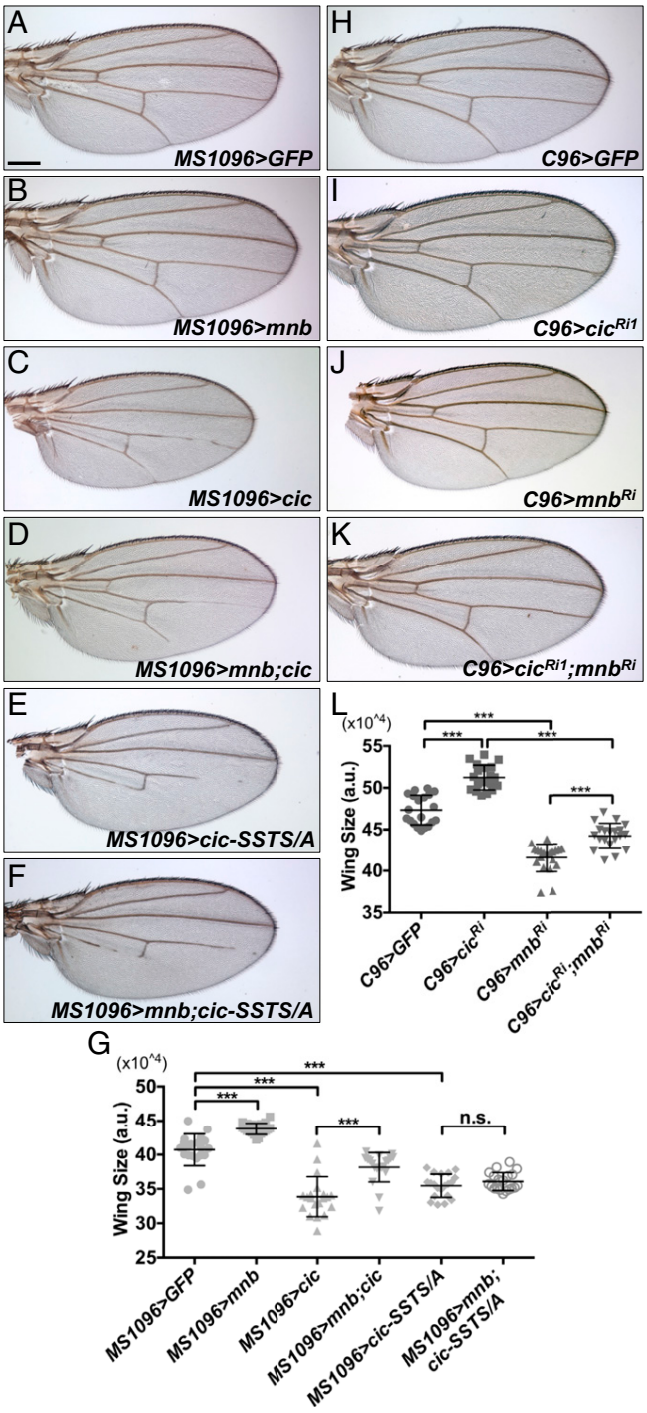


Fig. 4. Mnb opposes Cic function in controlling wing growth. (*A–F*) Wings from adult female flies expressing *UAS-GFP* as a control (*A*), *UAS-mnb* (*B*), *UAS-cic* (*C*), *UAS-mnb* together with *UAS-cic* (*D*), *UAS-cic-SSTS/A* (*E*), and *UAS-mnb* together with *UAS-cic-SSTS/A* (*F*) using the *MS1096-GAL4* driver. (*G*) Quantification of the wing areas for the genotypes shown in *A–F* ($n = 20$ for each genotype). (*H–K*) Wings from adult female flies expressing *UAS-GFP* as a control (*H*), $UAS\text{-}cic^{RNAi1}$ (*I*), $UAS\text{-}mnb^{RNAi}$ (*J*), and $UAS\text{-}cic^{RNAi1}$ together with $UAS\text{-}mnb^{RNAi}$ (*K*) using the *C96-GAL4* driver. (*L*) Quantification of the wing areas for the genotypes shown in *H–K* ($n = 20$ for each genotype). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Statistical significance was analyzed by using Student's *t* test. Error bars represent SD. (Scale bar: 200 μm.)

Mnb Opposes Cic Function in Controlling Wing and Eye Growth. We next investigated whether the inhibitory effects of Mnb/Wap on Cic were involved in the control of organ growth. Overexpression of Mnb using the wing pouch *MS1096-GAL4* driver (31) promoted wing growth (Fig. 4 *B* and *G*). Conversely, overexpression of Cic or the Cic-SSTS/A mutant resulted in a reduction of wing size (Fig. 4 *C*, *E*, and *G*). Whereas coexpression of Mnb with Cic suppressed the smaller wing size associated with Cic overexpression (Fig. 4 *C*, *D*, and *G*), coexpression of Mnb with Cic-SSTS/A did not modify this phenotype (Fig. 4 *E–G*). These data suggest that Mnb regulates Cic function at least in part through the phosphorylation of the SSTS residues. We also asked whether *mnb* and *cic* would display opposing effects on growth in a reduction-of-function context. RNAi depletion of *cic* using the *MS1096-GAL4* driver caused a severe defect in wing development. We thus used a weaker driver, *C96-GAL4*, which is expressed primarily around the wing margin (32), to study the effects of reduced levels of *mnb* and *cic*. Knockdown of *cic* caused wing overgrowth (Fig. 4 *I* and *L*), and RNAi depletion of *mnb* resulted in an opposite effect (Fig. 4 *J* and *L*). Importantly, RNAi depletion of *cic* partially rescued the small wing phenotype induced by expression of *mnb-RNAi* (Fig. 4 *K* and *J*). Mutually antagonistic effects of Mnb and Cic on growth were also observed in the eye (Fig. S6). Collectively, we conclude that Mnb and Wap promote wing and eye growth by antagonizing the growth-restricting function of Cic.

Reduction of *cic* Level Restores Adult Brain Size in *mnb* Mutants. Mnb was originally identified in a genetic screen for mutants with altered brain structure (19). Mutant *mnb* adult animals have smaller brains, with the optic lobes (OLs) most significantly affected (19). In *Drosophila* development, the size of the central brain (CB) is determined by the proliferative ability of the neuroblasts (NBs) that are of embryonic origin, whereas the OLs are generated by the neuroepithelium (NE), which gives rise to the OL NBs during the larval stages (33). To identify the tissue origins of the reduction in adult brain size, we studied the larval and pupal brains from the wild-type and mnb^{d419} animals (mnb^{d419} is a null allele; ref. 34). The volumes of the larval and pupal brains in the mnb^{d419} mutants were significantly smaller than controls (Fig. S7), suggesting that the effects of loss of *mnb* can be traced to these developmental stages. We asked whether the smaller OLs in *mnb* mutants could result from altered proliferation in the NE and/or NB regions during the larval stages. The widths of both the NB and NE regions in the larval brains from mnb^{d419} animals were significantly reduced, compared with controls (Fig. 5 *A–D*, *G*, and *H*). Conversely, overexpression of Mnb in MARCM clones resulted in an increase of the width of NE specifically in the clone area (Fig. 5 *E* and *F*). These results suggest that Mnb is required for the proper growth of both the NE and NB regions in the OL. Additionally, Mnb may be involved in controlling the timing of NE to NB differentiation (35).

We next asked whether interactions among *cic*, *mnb*, and *wap* were involved in the control of adult brain size. RNAi knockdown of *mnb* or *wap* with a ubiquitous *da-GAL4* driver (36) resulted in a smaller adult brain, especially in the optic lobes (Fig. 5 *K*, *L*, and *O*). This result suggests that both Mnb and Wap are required for normal brain growth. Knockdown of *cic* resulted in an increased adult brain size (Fig. 5 *J* and *O*). Strikingly, depletion of *cic* strongly suppressed the small brain phenotype caused by the knockdown of *mnb* (Fig. 5 *K*, *M*, and *O*), suggesting that Mnb promotes brain growth via down-regulation of Cic. Similarly, RNAi depletion of *cic* rescued the smaller brain phenotype of *wap-RNAi* (Fig. 5 *L*, *N*, and *O*). Overall, our results implicate Cic, Mnb, and Wap in a common pathway controlling organ growth and suggest that at least some of the growth-promoting functions of Mnb and Wap are mediated via their inhibition of Cic activity.

Mnb and ERK Have Additive Effects on Cic Activity. Our results so far have shown that Mnb is required for inhibiting Cic activity in various tissue contexts, which is also how ERK transmits signals

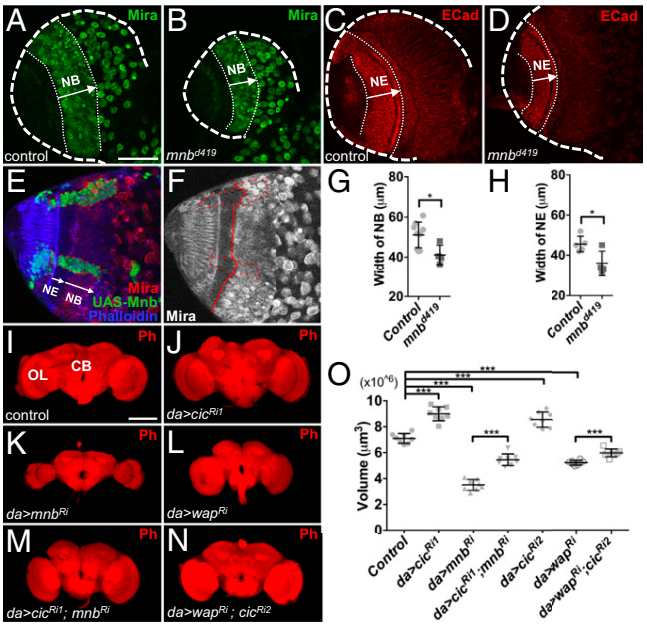


Fig. 5. Reduction in *cic* level restores adult brain size in *mnb* mutants. (*A* and *B*) Neuroblast (NB) regions (Mira-positive cells) in larval CNS from control (w^{1118}) (*A*) and mnb^{d419} animals (*B*). (*C* and *D*) Neuroepithelium (NE) regions (E-cad positive cells) in larval CNS from control (w^{1118}) (*C*) and mnb^{d419} animals (*D*). (*E* and *F*) NE region is expanded cell-autonomously in *UAS-mnb* overexpression clones (marked in green in *E*). Dotted red line, clone areas; solid red line, boundary between NB and NE. (*G*) Quantification of results in *A* and *B* ($n = 9, 4$). (*H*) Quantification of results in *C* and *D* ($n = 5, 4$). (*I*–*N*) Brains from adult female flies with the indicated genotypes. *da-GAL4* driver was used to drive the expression of *UAS-GFP* (*I*), *UAS-*cic^{RNAi1} (*J*), *UAS-*mnb^{RNAi} (*K*), *UAS-*wap^{RNAi} (*L*), *UAS-*cic^{RNAi1} together with *UAS-*mnb^{RNAi} (*M*), or *UAS-*cic^{RNAi2} together with *UAS-*wap^{RNAi} (*N*). Ph, phalloidin stain. (*O*) Quantification of brain volumes for the genotypes shown in *I*–*N* ($n = 8$ for each genotype). (Scale bars: *A*–*H*, 50 μm; *I*–*N*, 100 μm.) $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Error bars represent SD. Statistical significance was analyzed by using Student's *t* test.

Discussion

Our knowledge of upstream signals controlling Cic activity has been largely limited to its regulation by the RTK/ERK pathway (2). This study identifies a previously unknown mechanism for the regulation of Cic by the kinase Mnb and its adaptor Wap. Wap facilitates Mnb-dependent phosphorylation of Cic in the amino-terminal region, which is necessary for down-regulation of Cic activity. We found that the primary mechanism of Cic down-regulation by Mnb is through the relief of Cic-dependent transcriptional repression. Given that the DYRK family kinases autoactivate themselves soon after translation (37), it is likely that the effects of Mnb and Wap on Cic are constitutive.

Inhibition of Cic activity by Mnb/Wap has two developmentally important consequences (Fig. 6*E*). First, this regulation is important for the proper growth of several organs, such as the wings, eyes, and the brain. Second, down-regulation of Cic activity by Mnb/Wap is required for proper tissue patterning. Given the broad expression patterns of Cic, Mnb, and Wap (Fig. S2), the inhibitory mechanism we describe appears to operate in most, if not all, cells. In relation to ERK, the contribution from Mnb and Wap to Cic down-regulation depends on the tissue context and includes three possible scenarios: In some cells (e.g., developing vein L5 in the wing), both pathways are required for complete inhibition of Cic and operate additively. In other cells, ERK is the primary inhibitory signal, whereas the contribution of Mnb/Wap is less prominent (e.g., veins L2 and L3). Finally, in yet other cells in which ERK is not active, the function of Mnb and Wap to limit Cic activity would be dominant.

In addition to the RTK/ERK pathway, Cic was also shown to be regulated by Hippo signaling (18), and we have previously implicated Mnb and Wap as Hippo pathway regulators downstream of Dachsous (20). In this study, we found that knockdown of *sd*, a required component of Hippo signaling, did not affect the pattern of expression of *CUASC-LacZ*, and that knockdown of *mnb* or *wap* did not alter the pattern of ERK activation (Fig. S5), suggesting that Mnb and Wap control Cic activity independently from ERK and Hippo signaling. Altogether, current evidence suggests that Cic functions as an integrator of upstream developmental signals from RTKs to control growth and patterning. We next asked whether the effects of ERK and Mnb on Cic are additive by first individually and then simultaneously reducing their ability to inhibit Cic. ERK-mediated down-regulation of Cic depends on the conserved C2 motif located in region Cic3, which serves as the ERK docking site (5). Deletion of the C2 motif abrogated Cic–ERK interaction (5), and a single amino acid substitution, F1054A, in the C2 motif (QQFILAPTPAQLG) reduced the binding of Cic to ERK (26). Importantly, deletion of the C2 domain did not affect the binding of Cic to Mnb (Fig. S8). Using CRISPR/Cas9-mediated mutagenesis, we generated a *cic* allele (cic^3) lacking residue F1054, which is predicted to specifically disrupt the interaction of Cic with ERK. Most of the cic^3 mutant animals showed normal wing vein pattern (Fig. 6 *A* and *B*); however, a partial loss of vein L5 was observed in ~30% of adult flies, indicating that this is a gain-of-function mutation. RNAi depletion of *mnb* using *C5-GAL4* resulted in a partial loss of veins L4 and L5 (Fig. 6*C*, arrowheads), which is in agreement with our observation that *CUASC-lacZ* expression was lost in the L5 region in this background (Fig. 3*G*). We reasoned that if Mnb and ERK had additive effects on Cic activity, reduction of *mnb* level in the cic^3 background would cause a more severe vein loss phenotype compared with depletion of *mnb* alone. Indeed, we observed not only a more severe loss of veins L4 and L5, but also partial loss of veins L2 and L3 in the *C5 > mnbRi;* cic^3 animals (Fig. 6*D*). We conclude that Mnb and ERK function additively to regulate wing tissue patterning via inhibition of Cic activity.

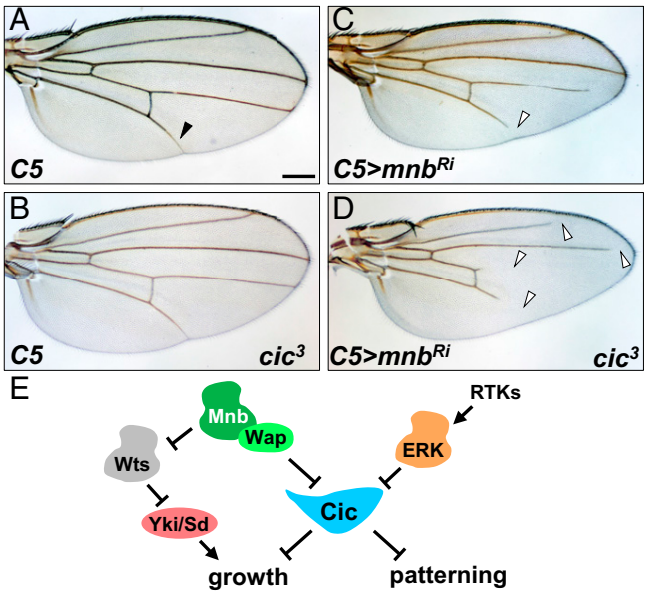


Fig. 6. Mnb and ERK function additively to inhibit Cic. Wings from adult female flies of the following genotypes: *C5-GAL4* (*A*), *C5-GAL4* cic^3/+ (*B*), *C5-GAL4/UAS-*mnb^{RNAi} (*C*), *C5-GAL4* cic^3/*UAS-*mnb^{RNAi} (*D*). (Scale bar: 200 μm.) (*E*) Cic integrates upstream signals to control organ growth and tissue patterning.

that converge on Cic to limit its activity, which is necessary for the proper execution of developmental programs responsible for tissue patterning and organ growth (Fig. 6*E*). Cic controls these developmental programs by direct binding to the enhancers of the genes encoding regulators of tissue patterning and cell proliferation in *Drosophila* and mammals (6, 8, 38, 39).

Interactions between Mnb/Wap and Cic in the brain have interesting parallels in human biology. The majority (>70%) of oligodendrogliomas, which are aggressive brain tumors, have been recently shown to harbor loss-of-function mutations in CIC, suggesting that it functions as a tumor suppressor (11, 12). Higher expression of DYRK1A was also found in a subset of oligodendroglioma patient samples (40), raising a possibility that DYRK1A may suppress Cic activity in human cells, much like Mnb does in *Drosophila*. A connection between DYRK1A and Cic in controlling brain development may extend even deeper, because both proteins have been implicated in neurodegenerative diseases (9, 41–43).

Materials and Methods

Experimental procedures are provided in *SI Materials and Methods*. They include a description of *Drosophila* stocks; information on antibodies, expression plasmids, and cell culture; procedures for luciferase reporter assays and mass spectrometry; and methods used for quantification of wing and brain size.

ACKNOWLEDGMENTS. We thank Iswar Hariharan for the Cic antibody; the Bloomington *Drosophila* Stock Center, Vienna *Drosophila* Resource Centre, and Developmental Studies Hybridoma Bank for their services; Walter Flores for help with preparation of Cic mutant constructs; and Jonathan Celli for help with brain size quantification. Mass spectrometry was performed at the Taplin Mass Spectrometry Facility at Harvard Medical School. This work was supported by the NIH Grants GM105813 (to A.V. and Ken Moberg) and HD085870 (to A.V. and S.Y.S.), L.Y. was supported by the University of Massachusetts Boston Sanofi Genzyme Doctoral Fellowship, S.Y.S. was supported by the NIH Grant GM086537, and G.J. is supported by Spanish Government Grant BFU2014-52863-P and Fundació La Marató de TV3 Grant 20131730.

1. Jiménez G, Guichet A, Ephrussi A, Casanova J (2000) Relief of gene repression by torso RTK signaling: Role of capicua in Drosophila terminal and dorsoventral patterning. *Genes Dev* 14(2):224–231.
2. Jiménez G, Shvartsman SY, Paroush Z (2012) The Capicua repressor–a general sensor of RTK signaling in development and disease. *J Cell Sci* 125(Pt 6):1383–1391.
3. Goff DJ, Nilson LA, Morisato D (2001) Establishment of dorsal-ventral polarity of the Drosophila egg requires capicua action in ovarian follicle cells. *Development* 128(22): 4553–4562.
4. Roch F, Jiménez G, Casanova J (2002) EGFR signalling inhibits Capicua-dependent repression during specification of Drosophila wing veins. *Development* 129(4): 993–1002.
5. Astigarraga S, et al. (2007) A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. *EMBO J* 26(3):668–677.
6. Ajuria L, et al. (2011) Capicua DNA-binding sites are general response elements for RTK signaling in Drosophila. *Development* 138(5):915–924.
7. Tseng AS, et al. (2007) Capicua regulates cell proliferation downstream of the receptor tyrosine kinase/ras signaling pathway. *Curr Biol* 17(8):728–733.
8. Jin Y, et al. (2015) EGFR/Ras signaling controls Drosophila intestinal stem cell proliferation via Capicua-regulated genes. *PLoS Genet* 11(12):e1005634.
9. Lam YC, et al. (2006) ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell* 127(7):1335–1347.
10. Bettegowda C, et al. (2011) Mutations in CIC and FUBP1 contribute to human oligodendroglioma. *Science* 333(6048):1453–1455.
11. Sahm F, et al. (2012) CIC and FUBP1 mutations in oligodendrogliomas, oligoastrocytomas and astrocytomas. *Acta Neuropathol* 123(6):853–860.
12. Wesseling P, van den Bent M, Perry A (2015) Oligodendroglioma: Pathology, molecular mechanisms and markers. *Acta Neuropathol* 129(6):809–827.
13. Grimm O, et al. (2012) Torso RTK controls Capicua degradation by changing its subcellular localization. *Development* 139(21):3962–3968.
14. Kim Y, et al. (2010) MAPK substrate competition integrates patterning signals in the Drosophila embryo. *Curr Biol* 20(5):446–451.
15. Kim Y, et al. (2011) Substrate-dependent control of MAPK phosphorylation in vivo. *Mol Syst Biol* 7(7):467.
16. Dissanayake K, et al. (2011) ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription factors via the transcriptional repressor capicúa. *Biochem J* 433(3):515–525.
17. Lim B, et al. (2013) Kinetics of gene derepression by ERK signaling. *Proc Natl Acad Sci USA* 110(25):10330–10335.
18. Herranz H, Hong X, Cohen SM (2012) Mutual repression by bantam miRNA and Capicua links the EGFR/MAPK and Hippo pathways in growth control. *Curr Biol* 22(8): 651–657.
19. Tejedor F, et al. (1995) minibrain: A new protein kinase family involved in postembryonic neurogenesis in Drosophila. *Neuron* 14(2):287–301.
20. Degoutin JL, et al. (2013) Riquiqui and minibrain are regulators of the hippo pathway downstream of Dachsous. *Nat Cell Biol* 15(10):1176–1185.
21. Morriss GR, et al. (2013) The Drosophila wings apart gene anchors a novel, evolutionarily conserved pathway of neuromuscular development. *Genetics* 195(3): 927–940.
22. Veraksa A (2013) Regulation of developmental processes: Insights from mass spectrometry-based proteomics. *Wiley Interdiscip Rev Dev Biol* 2(5):723–734.
23. Skurat AV, Dietrich AD (2004) Phosphorylation of Ser640 in muscle glycogen synthase by DYRK family protein kinases. *J Biol Chem* 279(4):2490–2498.
24. Himpel S, et al. (2000) Specificity determinants of substrate recognition by the protein kinase DYRK1A. *J Biol Chem* 275(4):2431–2438.
25. Zhai B, Villén J, Beausoleil SA, Mintseris J, Gygi SP (2008) Phosphoproteome analysis of Drosophila melanogaster embryos. *J Proteome Res* 7(4):1675–1682.
26. Futran AS, Kyin S, Shvartsman SY, Link AJ (2015) Mapping the binding interface of ERK and transcriptional repressor Capicua using photocrosslinking. *Proc Natl Acad Sci USA* 112(28):8590–8595.
27. Brunner D, et al. (1994) A gain-of-function mutation in Drosophila MAP kinase activates multiple receptor tyrosine kinase signaling pathways. *Cell* 76(5):875–888.
28. Bosch JA, Tran NH, Hariharan IK (2015) CoinFLP: A system for efficient mosaic screening and for visualizing clonal boundaries in Drosophila. *Development* 142(3): 597–606.
29. Yeh E, Gustafson K, Boulianne GL (1995) Green fluorescent protein as a vital marker and reporter of gene expression in Drosophila. *Proc Natl Acad Sci USA* 92(15): 7036–7040.
30. Zhang L, et al. (2008) The TEAD/TEF family of transcription factor Scalloped mediates Hippo signaling in organ size control. *Dev Cell* 14(3):377–387.
31. Capdevila J, Guerrero I (1994) Targeted expression of the signaling molecule decapentaplegic induces pattern duplications and growth alterations in Drosophila wings. *EMBO J* 13(19):4459–4468.
32. Helms W, et al. (1999) Engineered truncations in the Drosophila mastermind protein disrupt Notch pathway function. *Dev Biol* 215(2):358–374.
33. Sousa-Nunes R, Cheng LY, Gould AP (2010) Regulating neural proliferation in the Drosophila CNS. *Curr Opin Neurobiol* 20(1):50–57.
34. Hong SH, et al. (2012) Minibrain/Dyrk1a regulates food intake through the Sir2-FOXO-sNPF/NPY pathway in Drosophila and mammals. *PLoS Genet* 8(8):e1002857.
35. Reddy BV, Rauskolb C, Irvine KD (2010) Influence of fat-hippo and notch signaling on the proliferation and differentiation of Drosophila optic neuroepithelia. *Development* 137(14):2397–2408.
36. Wodarz A, Hinz U, Engelbert M, Knust E (1995) Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of Drosophila. *Cell* 82(1):67–76.
37. Becker W, Sippl W (2011) Activation, regulation, and inhibition of DYRK1A. *FEBS J* 278(2):246–256.
38. Kawamura-Saito M, et al. (2006) Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. *Hum Mol Genet* 15(13):2125–2137.
39. Forés M, et al. (2015) Origins of context-dependent gene repression by capicua. *PLoS Genet* 11(1):e1004902.
40. Pozo N, et al. (2013) Inhibition of DYRK1A destabilizes EGFR and reduces EGFR-dependent glioblastoma growth. *J Clin Invest* 123(6):2475–2487.
41. Guimerá J, et al. (1996) A human homologue of Drosophila minibrain (MNB) is expressed in the neuronal regions affected in Down syndrome and maps to the critical region. *Hum Mol Genet* 5(9):1305–1310.
42. Møller RS, et al. (2008) Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. *Am J Hum Genet* 82(5):1165–1170.
43. Tejedor FJ, Hämmerle B (2011) MNB/DYRK1A as a multiple regulator of neuronal development. *FEBS J* 278(2):223–235.
44. Poon CL, Mitchell KA, Kondo S, Cheng LY, Harvey KF (2016) The Hippo pathway regulates neuroblasts and brain size in Drosophila melanogaster. *Curr Biol* 26(8): 1034–1042.
45. Shaner NC, et al. (2008) Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat Methods* 5(6):545–551.
46. Port F, Chen HM, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc Natl Acad Sci USA* 111(29):E2967–E2976.
47. Kyriakakis P, Tipping M, Abed L, Veraksa A (2008) Tandem affinity purification in Drosophila: The advantages of the GS-TAP system. *Fly (Austin)* 2(4):229–235.
48. Frolov MV, et al. (2001) Functional antagonism between E2F family members. *Genes Dev* 15(16):2146–2160.
49. Neumüller RA, et al. (2012) Stringent analysis of gene function and protein-protein interactions using fluorescently tagged genes. *Genetics* 190(3):931–940.
50. Choi H, et al. (2011) SAINT: Probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* 8(1):70–73.

Supporting Information

Yang et al. 10.1073/pnas.1609417113

SI Materials and Methods

***Drosophila melanogaster* Stocks.** All *Drosophila* stocks were maintained on standard yeast-cornmeal-agar medium at 25 °C or 18 °C as indicated. *MS1096-GAL4*, *da-GAL4*, *C96-GAL4*, *en-GAL4, hh-GAL4 UAS-GFP*, ey-*FLP UAS-dcr2* (#58757), and *CoinFLP UAS-GFP* (#58751) were from the Bloomington *Drosophila* Stock Center. *UAS-wap RNAi* (KK 107076), *UAS-mnb RNAi* (GD 28628), *UAS-cic RNAi1* (KK 103012), *UAS-cic RNAi2* (GD 40867), and *UAS-sd RNAi* (KK 101497) were from the Vienna *Drosophila* Resource Center (VDRC). *UAS-cic*, *UAS-mnb*, and *UAS-wap* were made by using standard cloning methods, and transgenic lines were generated by Genetic Services. *Cic-Venus* is a genomic Cic rescue construct from ref. 13. Other stocks were *C5-GAL4* and *CUASC-lacZ* (6), mnb^{d419} (34). *mnb-tagRFP-T* and *wap-Venus* were generated by CRISPR/Cas9-mediated targeted transgene integration as described (44). The targeting vectors comprised the tagRFP-T (45) or Venus gene, the 3xP3-RFP gene for transformant selection and 1-kb homology arms. The vectors were designed such that the fluorescent protein gene is inserted immediately in front of the stop codon of the target gene and expressed as a fusion protein. The cic^3 allele was obtained by CRISPR/Cas9-mediated mutagenesis. A guide RNA (gRNA) sequence (5′-GTGGGTGCCAAAATAAACTGC-3′) targeting the C2 coding sequence was subcloned in vector *pCFD3* (46) and inserted at the *attP40* genomic site via PhiC31-based integration. Transgenic gRNA males were crossed to *nanos-cas9* females, and the resulting founder males were then crossed to *TM3*-bearing females for recovery of mutations. Induced alleles were identified by sequencing PCR products amplified from candidate flies.

Immunohistochemistry and Immunoblotting. Primary antibodies were as follows: mouse anti-β-galactosidase (LacZ) 1:100 (Promega), mouse anti-dpERK 1:100 (Sigma), rabbit anti-GFP 1:1,000 (Abcam), mouse anti-V5 1:1,000 (Sigma), rabbit anti-Flag 1:1,000 (Sigma), rabbit anti-HA 1:1,000 (Sigma), guinea pig anti-Cic 1:100 (gift from Iswar Hariharan, University of California, Berkeley, CA), rabbit anti-DCAF7 1:100 (Novus Biologicals), rabbit anti-tRFP 1:1,000 (Evrogen), mouse anti-Mira 1:100 (gift from Alex Gould, The Francis Crick Institute, London), mouse anti-Ecad 1:50 (Developmental Studies Hybridoma Bank), and mouse anti-Engrailed 1:5 (Developmental Studies Hybridoma Bank). Secondary antibodies used were as follows: Alexa Donkey anti Rabbit-647 (Life Technologies), Alexa Donkey anti-Mouse 488 (Abcam), Alexa Donkey anti-Rabbit 555 (Life Technologies), IRDye 800CW Donkey anti-Rabbit IgG (LI-COR), IRDye 680CW Donkey anti-Guinea pig IgG (LI-COR), and IRDye 680CW Donkey anti-Mouse IgG (LI-COR). Dissected imaginal discs were stained as in ref. 20. Stained tissues were mounted with Prolong Gold anti-fade mounting reagent with DAPI (Life Technologies), and images were acquired with Zeiss LSM 510 or Zeiss LSM 880 confocal microscopes. Tissues from the *Mnb-tRFP, Wap-Venus*, and *Cic-Venus* stocks were mounted in 90% (vol/vol) glycerol, and images were acquired with the Nikon C2 confocal microscope and processed in Fiji.

Quantification of Wing and Brain Size. Wing area from 20 flies was quantified by using Adobe Photoshop. For quantification of brain size, flies were raised at 18 °C. Brains from eight 10-day-old female flies were dissected in cold 1× PBS, fixed with 4% formaldehyde/PBS, and stained with Alexa Fluor 594 Phalloidin (Life Technologies) overnight. Stained brains were mounted with Prolong Gold anti-fade mounting reagent with DAPI (Life Technologies). Brain volume measurements were done from 3D reconstructions of 2.5 μm-spaced confocal Z stacks acquired with a Zeiss LSM 880 or Leica SP5 confocal microscopes, using a 3D Viewer plugin in Fiji or Volocity software. Significance was calculated with a Student's *t* test. In all figures, the indications used are the following: $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

Expression Plasmids and Cell Culture. For establishing a stable S2 cell line, full-length Cic ORF was tagged with SBP at the C terminus and cloned into pMK33 vector (47). Full-length ORF and fragments of Cic were cloned into pMT/V5-His A vector (Invitrogen). Full-length Wap was tagged with the Flag tag and cloned into the pMT vector. pMT-Mnb-HA and pMT-MnbKR-HA plasmids were from ref. 20. The pGL2-CUASC-Luc ("*CUASC-Luc*") reporter was generated by PCR amplification of the promoter regions of pC4PLZ-CUASC vector (6) with primers CUASC-XhoI-UP1 (5′-ATCGCTCGAGGAATTCCCAGTTTATG-3′) and CUASC-XhoI-DN1 (5′-GCTACTCGAGTTATCACCCACGGCTCTGCTC-3′), which was subcloned into pGL2 basic (Promega). Full-length GAL4 was cloned into pMT/V5-His A vector (Invitrogen) to generate pMT-GAL4. pIE4-lacZ was described (48).

Drosophila S2 cells were maintained in standard Schneider's S2 medium with FBS (Gibco) at 25 °C, and transfections were performed by using Effectene transfection reagent (Qiagen). In some instances, cells were treated with 30 μg of dsRNA specific for *wap*, *mnb*, or *rl* (ERK) for 96 h. After 96 h, cells were transfected with indicated plasmids. $CuSO_4$ was added to culture media at a final concentration of 0.35 mM for inducing expression. Cells were lysed by using default lysis buffer (50 mM Tris pH 7.5, 125 mM NaCl, 5% glycerol, 0.2% IGEPAL, 1.5 mM $MgCl_2$, 1 mM DTT, 25 mM NaF, 1 mM Na_3VO_4, 1 mM EDTA and 2× Complete protease inhibitor, Roche). Clear cell lysates were incubated with anti-V5 beads (Sigma), streptavidin beads (Pierce), GFP-Trap, or RFP-Trap resin (ChromoTek) for 2 h at 4 °C. Beads were washed three times with lysis buffer, and protein complexes were eluted with SDS buffer. To visualize differences in Cic1 mobility, 6% Tris-Glycine gels were used with 50 μM Phos-tag (Wako Laboratory Chemicals) and 100 μM $MnCl_2$.

Luciferase Reporter Assays. Cells were cotransfected with the luciferase reporter vector pGL2-CUASC-Luc and with the effector plasmids. pIE4-LacZ was used to normalize transfection efficiencies. Each transfection point was assayed in triplicate, and each experiment was repeated three times. Luciferase and β-galactosidase activities were measured in S2 cell lysates by Luc-Screen Extended-Glow Luciferase Reporter Gene Assay System (Thermo Fisher) and Galacto-Star One-Step β-Galactosidase Reporter Gene Assay System (Thermo Fisher), respectively. Luminescence signals were acquired on POLARstar Omega multifunction microplate reader (BMG Labtech).

Mass Spectrometry. Cic-interacting proteins were purified from *Drosophila* S2 cells (Cic-SBP) essentially as described in ref. 47, using a modified single-step procedure for SBP-tagged Cic. For expression in S2 cells, pMK33-Cic-SBP construct was transfected by using Effectene transfection reagent (Qiagen), and stable cell lines were selected in the presence of 300 μg/mL hygromycin (Sigma). Cic-SBP purifications were performed in two biological replicates. To analyze Cic complexes in vivo, 0- to 16-h Cic-Venus embryos were collected and the proteins were extracted and purified on GFP-Trap resin (ChromoTek) as described in ref. 49. The Cic-Venus construct uses genomic regulatory sequences

of the *cic* gene and, therefore, expresses the Cic-Venus protein at endogenous levels (13). Functionality of this construct was previously confirmed in a rescue assay (13). Cic-Venus purifications were performed in three biological replicates. Protein complexes were analyzed by nanoLC-MS/MS as described in ref. 47 at the Taplin Mass Spectrometry Facility at Harvard Medical School. Identified Cic-interacting proteins were analyzed by the SAINT program (50). A complete mass spectrometry dataset is shown in Dataset S1. Interactions with SAINT scores >0.8 were considered significant.

To identify Cic residues that are phosphorylated by Mnb, Cic1-V5 and Wap-Flag were coexpressed in S2 cells either with Mnb-HA or MnbKR-HA, Cic1-V5 was purified on anti-V5 agarose resin, and its phosphorylation was analyzed by nanoLC-MS/MS.

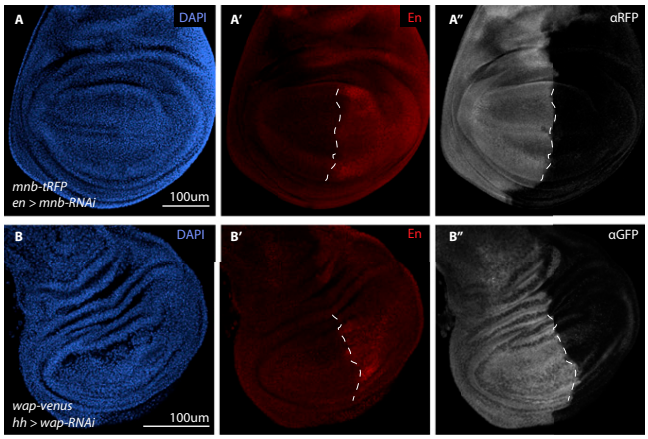


Fig. S1. Depletion of fluorescently tagged endogenous Mnb and Wap confirms the on-target effect of *mnb-RNAi* and *wap-RNAi*. (*A*–*B*″) Wing discs from third instar larvae of the indicated genotypes. The *en-GAL4* and *hh-GAL4* drivers are expressed in the posterior compartment of wing discs (red in *A*′ and *B*′) and were used to drive the expression of *UAS-mnb RNAi* (*A*–*A*″) and *UAS-wap RNAi* (*B*–*B*″) respectively. Cell nuclei were detected with DAPI (blue in *A* and *B*) whereas Mnb-tRFP (gray in *A*″) and Wap-Venus (gray in *B*″) were detected by anti-tRFP and anti-GFP antibodies, respectively. (Scale bars: 100 μm.) Dashed lines (*A*′ and *B*′) indicate the anterior-posterior compartment boundary of the wing pouch.

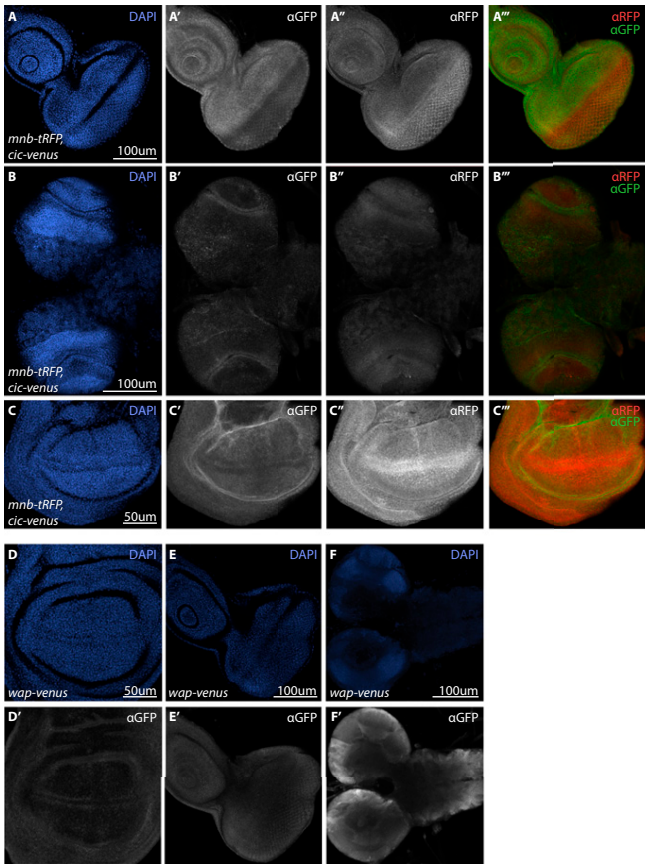


Fig. S2. Mnb-RFP, Cic-Venus, and Wap-Venus expression in third instar larval tissues. (*A*–*F*) Cic-Venus and Wap-Venus expression was detected by GFP antibody, Mnb-tRFP was detected by tRFP antibody and cell nuclei were detected by DAPI (blue). (*A*′–*C*‴) Cic-Venus and Mnb-tRFP expression in eye imaginal discs (*A*–*A*‴), brain optic lobes (*B*′–*B*‴), and the pouch region of wing imaginal discs (*C*′–*C*‴). For merged panels (*A*‴–*C*‴), Cic-Venus is in green and Mnb-tRFP is in red. (*D*–*F*′) Wap-Venus expression (gray) in wing imaginal discs (*D*′), eye imaginal discs (*E*′) and larval brain (*F*′). (Scale bars: *A*–*B*‴, *E*, and *F*, 100 μm; *C* and *D*, 50 μm.)

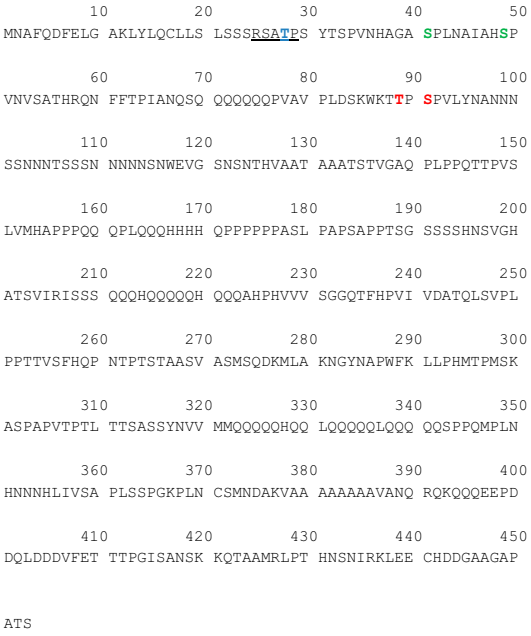


Fig. S3. Locations of Cic phosphorylation sites identified by mass spectrometry. A Cic1 region is shown (amino acids 1–453). The putative DYRK1A consensus is underlined, and the corresponding residue (T28) is highlighted in blue. T28 phosphorylation was not detected by mass spectrometry. S41 and S49 (green) were more highly phosphorylated in wild-type Mnb samples compared with MnbKR. T89 and S91 phosphorylations (red) were found exclusively in the wild-type Mnb samples.

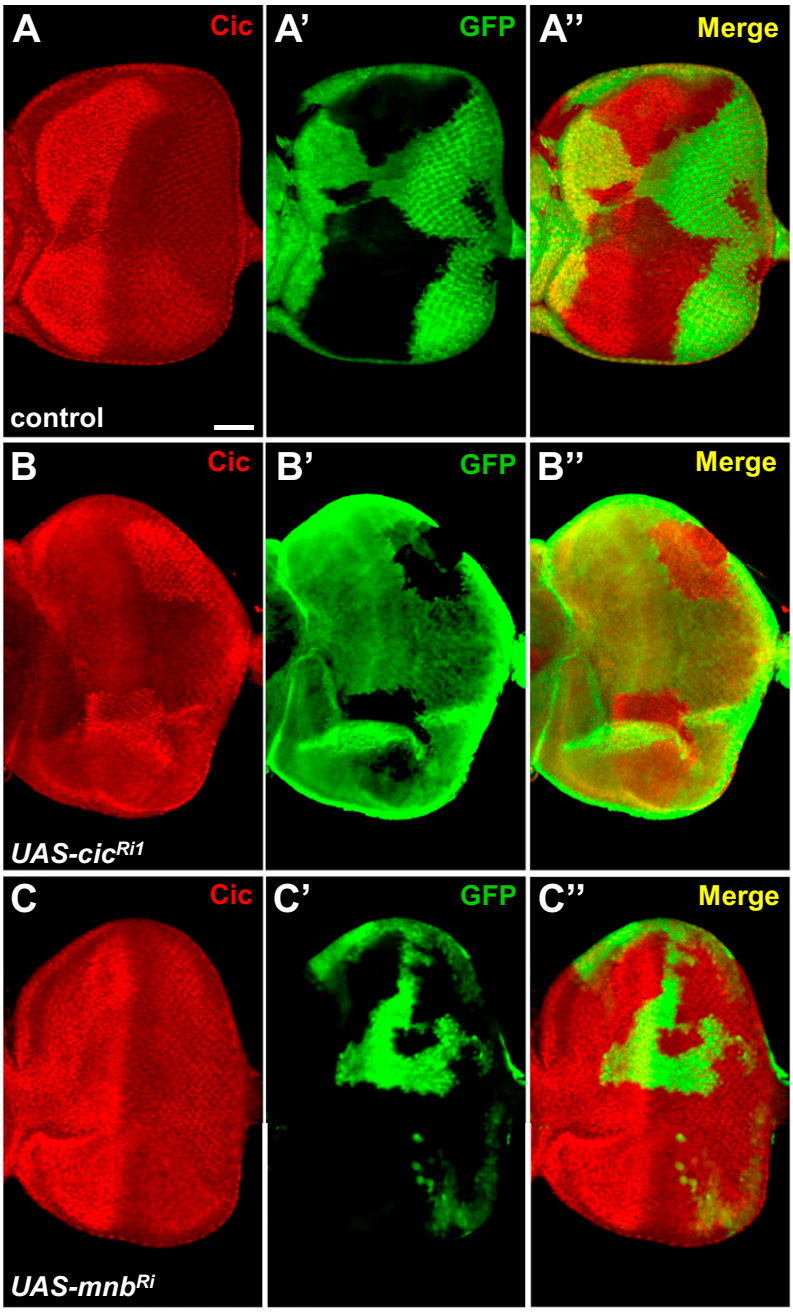


Fig. S4. RNAi depletion of Mnb does not increase Cic levels in imaginal eye discs. (*A*–*C*′′) Mosaic eye discs of the indicated genotypes generated by using CoinFLP-GAL4 and stained with anti-Cic antibody (red). GAL4-positive cells are marked with *UAS-GFP* (green). (*A*–*A*′′) Control clones do not affect Cic protein levels. (*B*–*B*′′) Knockdown of *cic* reduces Cic protein levels. (*C*–*C*′′) Knockdown of *mnb* does not increase Cic protein level or change Cic subcellular localization. (Scale bar: 50 μm.)

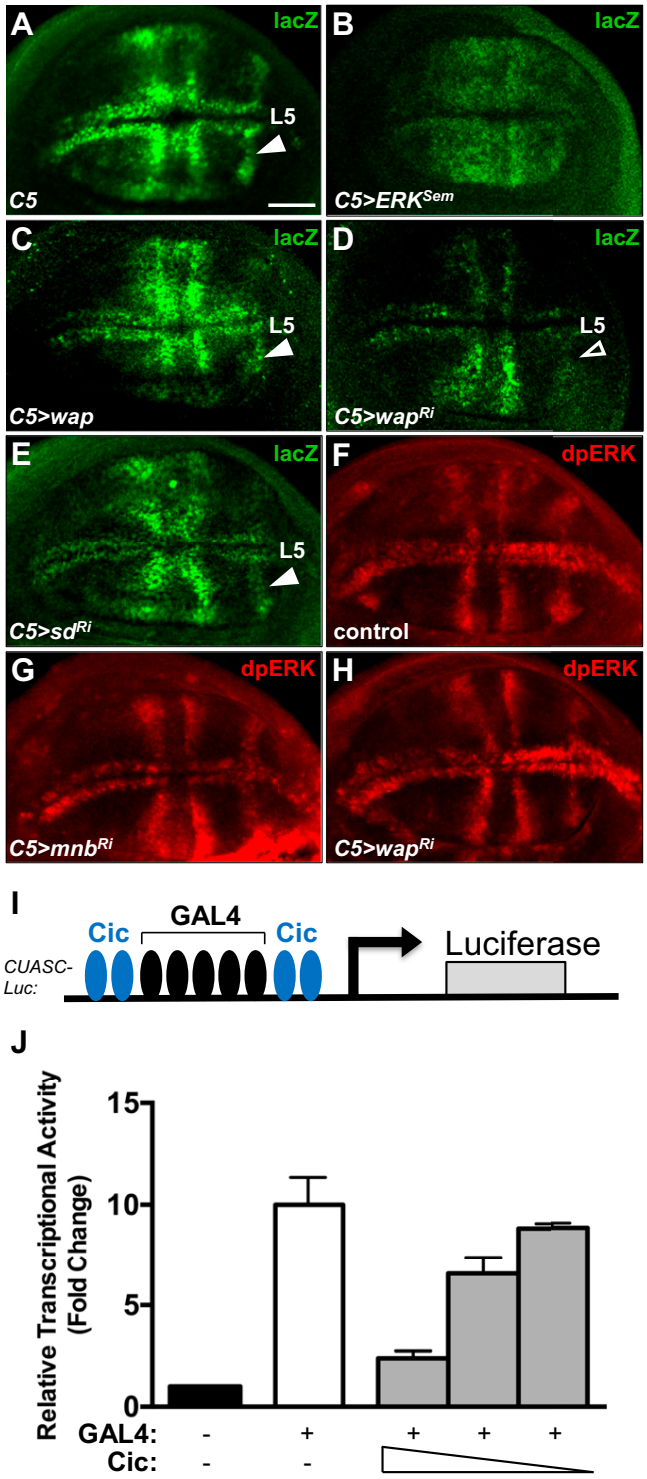


Fig. S5. Modulation of Cic repressor activity by Mnb and Wap does not require Sd and does not affect activity of ERK. (*A–E*) LacZ expression pattern resulting from *C5-GAL4*–directed activation of *CUASC-lacZ* in wing imaginal discs from control (*A*), *UAS-ERK*^{Sem} (*B*), *UAS-wap* (*C*), *UAS-wap*^{RNAi} (*D*), and *UAS-sd*^{RNAi} (*E*) larvae. Expression of ERK^{Sem} and Wap led to a broader activation of the reporter (*B* and *C*), whereas knockdown of *wap* resulted in lower LacZ expession, particularly in presumptive vein L5 (*D*). Knockdown of *sd* did not alter the normal pattern of expression in presumptive veins. (*F–H*) dpERK expression pattern in wing discs from control (*F*), *C5-GAL4* > *UAS-mnb*^{RNAi} (*G*), and *C5-GAL4* > *UAS-wap*^{RNAi} (*H*) larvae. (*I*) Schematic diagram of the *CUASC-Luc* reporter construct. (*J*) Dose-dependent repression of *CUASC-Luc* expression by Cic. S2 cells were cotransfected with the reporter *CUASC-Luc*, pMT-GAL4, and decreasing amounts of Cic-expressing plasmid (500 ng, 250 ng, 125 ng). The values shown are fold changes over the negative control set at 1. (Scale bar: 50 μm.)

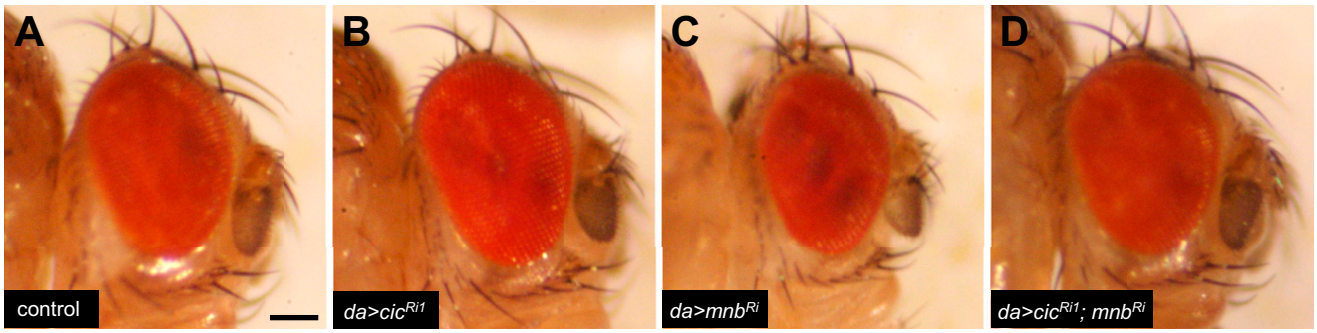


Fig. S6. Mnb opposes Cic function in controlling eye growth. Adult female flies expressing *UAS-GFP* (*A*), *UAS-cic*^{RNAi1} (*B*), *UAS-mnb-RNAi* (*C*), and *UAS-mnb-RNAi* together with *UAS-cic*^{RNAi1} (*D*) under the control of the *da-GAL4* driver. Knockdown of *mnb* results in a smaller eye (*C*), which is reversed by a concomitant knockdown of *cic* (*D*). (Scale bar: 100 μm.)

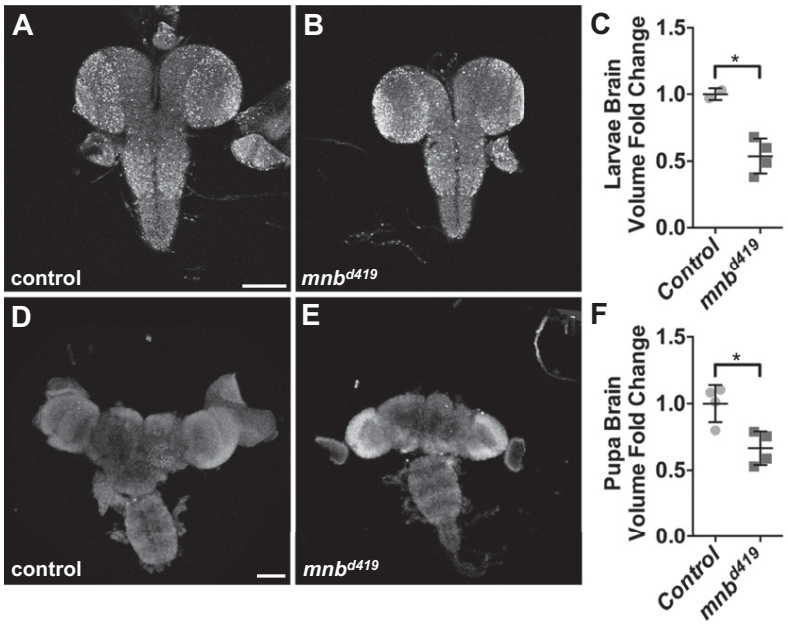


Fig. S7. Loss of *mnb* leads to smaller brain size in larvae and pupae. (A and B) Third instar larval brains from control (w^{1118}) (*A*) and mnb^{d419} (*B*) animals. (C) Quantification of the larval brain volumes in *A* and *B* ($n = 2, 4$). (*D* and *E*) Pupal brains from control (w^{1118}) (*D*) and mnb^{d419} (*E*) animals. (*F*) Quantification of the pupal brain volumes in *D* and *E* ($n = 4, 4$). *$P < 0.05$. Statistical significance was analyzed by using Student's *t* test. (Scale bars: 100 μm.)

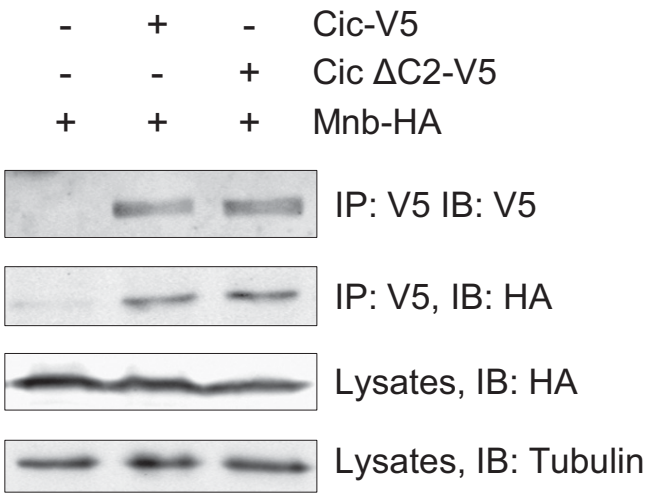


Fig. S8. Deletion of the C2 motif does not affect the binding between Cic and Mnb. Protein lysates from S2 cells transfected with the indicated plasmids were incubated with anti-V5 beads, and immunocomplexes were analyzed on Western blots probed with anti-V5, anti-HA, and anti-tubulin antibodies.

Dataset S1. Cic mass spectrometry data

Dataset S1

Cic-SBP purifications were performed from *Drosophila* S2 cells (two biological replicates), Cic-Venus purifications were performed from 0 to 16 h *Drosophila* embryos (three biological replicates). Peptide numbers identified by nanoLC-MS/MS were used for the analysis with the SAINT program. Cic is highlighted in blue; Wap and Mnb are highlighted in green.

PUBLICACIÓN 4: A new mode of DNA binding distinguishes Capicua from other HMG-box factors and explains its mutation patterns in cancer

Autores:

Marta Forés, Lucía Simón-Carrasco, Leiore Ajuria, Núria Samper, Sergio González-Crespo, Matthias Drosten, Mariano Barbacid, Gerardo Jiménez

Referencia:

PLoS Genet **13**, e1006622 (2017)

Resumen:

Las proteínas HMG-box, incluyendo los miembros de las familias Sox/SRY (Sox) y TCF/LEF1 (TCF), se unen a ADN a través de su HMG-box. Esta unión es relativamente débil y tanto los factores Sox como los TCF emplean distintos mecanismos para aumentar su afinidad y especificidad a ADN. En este artículo mostramos que Capicua (CIC), un represor transcripcional HMG-box implicado en la señalización Ras/MAPK y la progresión del cáncer, utiliza un modo distinto de unión a ADN que permite un reconocimiento selectivo de sus dianas. Hemos encontrado que, contrariamente a lo que se creía, la HMG-box de CIC no se une a ADN por sí sola, sino que requiere un motivo distante (llamado C1) presente en la región C-terminal de todas las proteínas CIC. Tanto el motivo HMG-box como el C1 son necesarios para la unión específica a lugares TGAATGAA o similares, no funcionan a través de dimerización, y son activos en ausencia de cofactores, sugiriendo que forman una estructura bipartita para la unión específica de secuencia a ADN. Demostramos que este mecanismo de unión opera a través del desarrollo y en células humanas, asegurando una regulación específica de las múltiples dianas de CIC. Parece que las proteínas HMG-box generalmente dependen de mecanismos auxiliares de unión a ADN para regular sus dianas apropiadas, pero cada sub-familia ha evolucionado estrategias únicas para este propósito. Finalmente, el papel clave del C1 en la unión a ADN también explica el hecho que este dominio sea un punto recurrente de mutaciones inactivantes en oligodendroglioma y otros tumores, mientras que debe ser preservado en las fusiones quiméricas oncogénicas CIC-DUX4 asociadas a sarcomas de tipo *Ewing-like*.

PLOS | GENETICS

RESEARCH ARTICLE

A new mode of DNA binding distinguishes Capicua from other HMG-box factors and explains its mutation patterns in cancer

Marta Forés[1], Lucía Simón-Carrasco[2], Leiore Ajuria[1], Núria Samper[1], Sergio González-Crespo[1], Matthias Drosten[2], Mariano Barbacid[2], Gerardo Jiménez[1,3]*

1 Institut de Biologia Molecular de Barcelona-CSIC, Barcelona, Spain, **2** Molecular Oncology Programme, Centro Nacional de Investigaciones Oncológicas, Madrid, Spain, **3** ICREA, Barcelona, Spain

* gjcbmc@ibmb.csic.es

Citation: Forés M, Simón-Carrasco L, Ajuria L, Samper N, González-Crespo S, Drosten M, et al. (2017) A new mode of DNA binding distinguishes Capicua from other HMG-box factors and explains its mutation patterns in cancer. PLoS Genet 13(3): e1006622. doi:10.1371/journal.pgen.1006622

Editor: Ken M. Cadigan, University of Michigan, UNITED STATES

Received: July 12, 2016

Accepted: February 8, 2017

Published: March 9, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by grants from the Spanish Ministry of Economy and Competitiveness (MINECO) (BFU2011-23611 and BFU2014-52863-P) to GJ; the European Research Council (ERC-AG/250297-RAS AHEAD), the EU-Framework Programme (LSHG-CT-2007-037665/CHEMORES, HEALTH-F2-2010-259770/LUNGTARGET and HEALTH-2010-260791/EUROCANPLATFORM), the

Abstract

HMG-box proteins, including Sox/SRY (Sox) and TCF/LEF1 (TCF) family members, bind DNA via their HMG-box. This binding, however, is relatively weak and both Sox and TCF factors employ distinct mechanisms for enhancing their affinity and specificity for DNA. Here we report that Capicua (CIC), an HMG-box transcriptional repressor involved in Ras/MAPK signaling and cancer progression, employs an additional distinct mode of DNA binding that enables selective recognition of its targets. We find that, contrary to previous assumptions, the HMG-box of CIC does not bind DNA alone but instead requires a distant motif (referred to as C1) present at the C-terminus of all CIC proteins. The HMG-box and C1 domains are both necessary for binding specific TGAATGAA-like sites, do not function via dimerization, and are active in the absence of cofactors, suggesting that they form a bipartite structure for sequence-specific binding to DNA. We demonstrate that this binding mechanism operates throughout *Drosophila* development and in human cells, ensuring specific regulation of multiple CIC targets. It thus appears that HMG-box proteins generally depend on auxiliary DNA binding mechanisms for regulating their appropriate genomic targets, but that each sub-family has evolved unique strategies for this purpose. Finally, the key role of C1 in DNA binding also explains the fact that this domain is a hotspot for inactivating mutations in oligodendroglioma and other tumors, while being preserved in oncogenic CIC-DUX4 fusion chimeras associated to Ewing-like sarcomas.

Author summary

Transcription factors bind specific sites in the genome via discrete protein domains that recognize their target DNA sequences. One such domain is the HMG-box, which is found in many chromatin and transcriptional regulators across species. Two salient groups of HMG-box proteins are the Sox/SRY and TCF/LEF1 factors, which are involved in multiple developmental and signaling processes. Extensive genetic and molecular studies have shown, however, that both groups of proteins do not simply bind DNA through their HMG-box, but rely either on additional protein domains or associated factors for

MINECO (SAF2011-30173 and SAF2014-59864-R), the Autonomous Community of Madrid (S2011/BDM-2470/ONCOCYCLE) to MB; and the Fundació La Marató de TV3 (20131730/1). LSC was supported by a fellowship from the Programa de Formación de Personal Investigador (FPI, MINECO). MB is a recipient of an Endowed Chair from the AXA Research Fund. GJ is an ICREA investigator. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

targeting their correct sites in the genome. In this work, we have focused on another HMG-box protein, Capicua (CIC), which has recently emerged as an important mediator of Ras/MAPK signaling in both *Drosophila* and mammals. We find that the HMG-box of CIC does not bind DNA alone and instead requires a separate conserved motif (C1) present in all CIC proteins. The C1 domain is restricted to CIC proteins and exhibits several properties that distinguish it from Sox and TCF domains involved in DNA binding. Thus, CIC proteins represent a separate sub-family of HMG-box factors that have evolved an independent mechanism for enhancing the DNA-binding capabilities of their HMG-box. Notably, our results also explain distinct patterns of human *CIC* mutations that either inactivate CIC tumor suppressor function or produce oncogenic fusions between CIC and the DUX4 activator factor.

Introduction

HMG-box factors are abundant nuclear proteins with highly diverse functions in the cell. They contain one or more HMG-box domains that bind the minor groove of DNA, bending the duplex away from the interaction site. Proteins with tandem HMG-box domains usually function as architectural and chromatin factors and do not exhibit DNA sequence specificity. In contrast, proteins with a single HMG-box domain, including Sox/SRY (Sox) and TCF/LEF1 (TCF) transcription factors, function as developmental regulators and bind specific AT-rich motifs in enhancers and promoters (reviewed in refs. [1–3]). In most cases, however, this binding is not sufficient for appropriate target selection. For example, Sox proteins rarely act on their own and are often assisted by partner factors that bind next to the Sox sites, thereby stabilizing the complex and providing the specificity needed for in vivo function [3]. Once tethered to DNA, HMG-box proteins can exert their transcriptional effects through additional interactions with co-activators or co-repressors.

The HMG-box protein Capicua (CIC) is a highly conserved transcriptional repressor distantly related to Sox and TCF factors [4]. Studies in *Drosophila* and mammals have shown that CIC controls multiple developmental decisions acting downstream of Receptor Tyrosine Kinase (RTK) signaling. In general, CIC represses RTK-responsive genes by binding to octameric TGAATGAA-like motifs in their promoters and enhancers, and this repression is relieved upon RTK-induced downregulation of CIC. In *Drosophila*, this mechanism controls anteroposterior and dorsoventral body patterning, intestinal stem cell proliferation, wing development, and other processes, providing a direct link between RTK activation and transcriptional derepression of CIC targets [5–14]. In mammals, CIC is similarly regulated by RTK signaling and controls essential processes such lung alveolarization and liver homeostasis [15–18]. Moreover, CIC has been implicated in distinct human pathologies including spinocerebellar ataxia type 1 [16,19] and various forms of cancer, particularly oligodendroglioma (OD) [20–23]. In cancer, CIC behaves mainly as a tumor and metastasis suppressor that is inactivated by somatic mutations [22–30], but it can also exert oncogenic effects resulting in Ewing-like sarcomas [31] (Fig 1). This latter role originates from chromosomal translocations where CIC becomes fused to a fragment of the DUX4 transcription factor [31–36]. CIC-DUX4 chimeras contain a nearly complete CIC sequence followed by the C-terminal portion of DUX4, which converts CIC into an activator and causes upregulation of CIC targets such as *ETV/PEA3* family genes [15,17,31].

Nevertheless, the mechanisms underlying CIC activity in normal and pathological processes are not well understood. One unresolved question concerns the role of a conserved

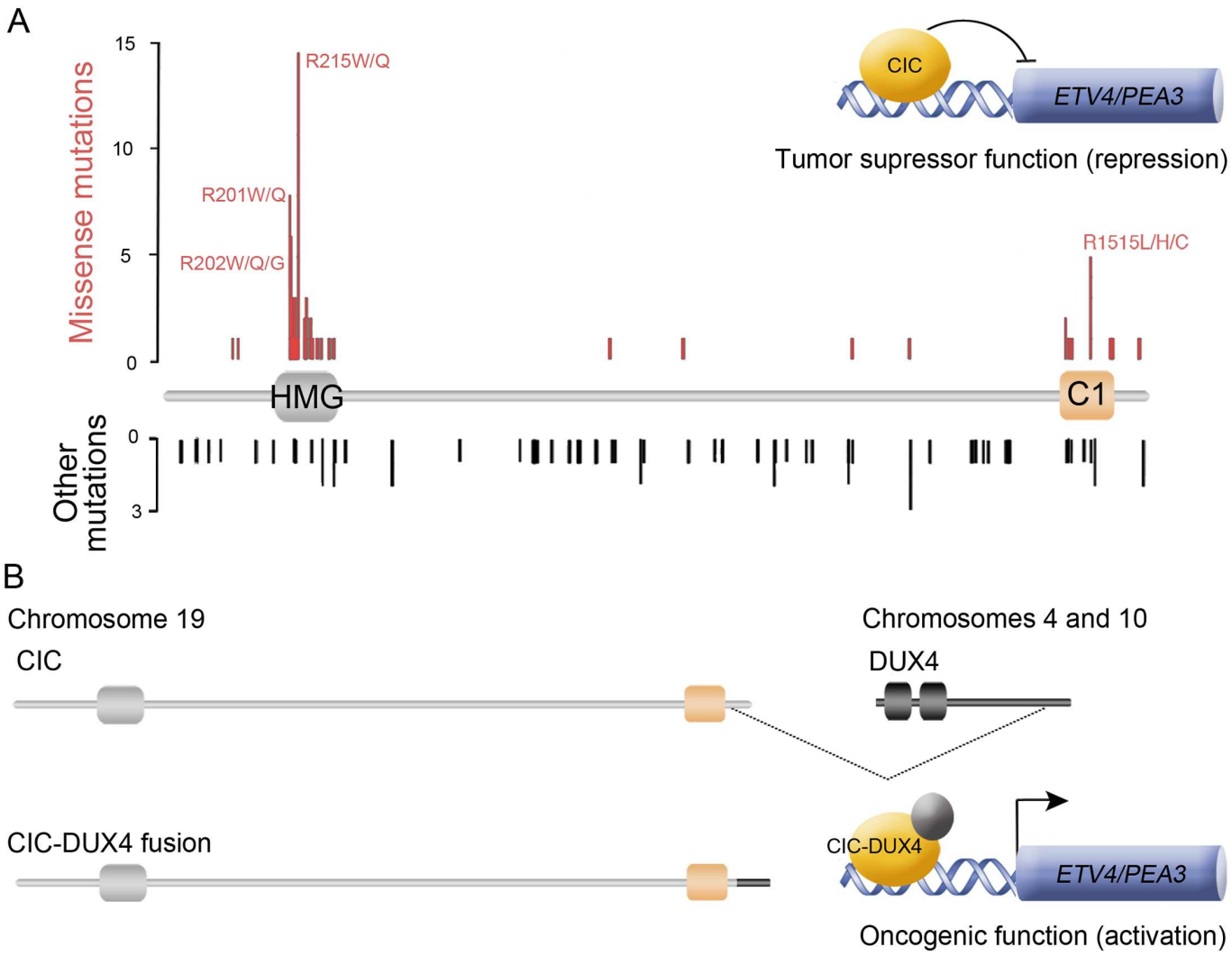


Fig 1. Patterns of *CIC* mutations in human OD and *CIC-DUX4* sarcomas. (A) Diagram of the CIC protein showing a set of curated mutations from the COSMIC database (http://cancer.sanger.ac.uk/cosmic). Only mutations corresponding to gliomas are shown. The tumor suppressor role of CIC in OD is thought to involve the repression of CIC targets such as the *ETV/PEA3* family genes [29, 30]. Note that missense mutations tend to cluster in the HMG-box and C1 domains. In contrast, nonsense and frameshift mutations (indicated as 'Other mutations') are distributed along the entire length of the protein, which is also consistent with a requirement for an intact C-terminal region. (B) Structure and function of oncogenic CIC-DUX4 fusions, which usually include most of the CIC protein (including the C1 domain) coupled to the C-terminal trans-activation domain of DUX4 [31,66]. The double homeodomain region of DUX4 is indicated by boxes.

doi:10.1371/journal.pgen.1006622.g001

domain present at the C-terminus of all CIC proteins. This domain (referred to as C1) does not resemble other known domains and appears to be functionally important. Thus, transgenic assays indicate that C1 is required for CIC repressor activity in early *Drosophila* embryos [9]. Also, systematic sequencing of human tumors has revealed multiple missense mutations mapping to the C1 sequence, arguing that C1 is essential for CIC function in suppressing growth and metastasis [22–27,30]. However, how C1 contributes to CIC function remains unknown.

In this work, we set out to investigate the mechanism of C1 action. Unexpectedly, we find that C1 plays a conserved essential role in CIC DNA binding activity. We show that neither

the HMG-box nor the C1 domain is capable of binding to DNA separately, but instead function together to mediate efficient DNA binding in both *Drosophila* and human cells. Thus, CIC employs a new mode of DNA binding that distinguishes it from Sox and TCF proteins, which lack the C1 domain and employ other mechanisms for enhancing their target specificity. Furthermore, our results explain the distinct patterns of human *CIC* mutations in OD and Ewing-like sarcomas, since the C1 domain should be required for the DNA binding activities of both CIC and CIC-DUX4 in those pathologies, respectively.

Results

The C1 domain is essential for multiple developmental functions in *Drosophila*

The C1 domain is highly conserved in all CIC orthologs across metazoans. The conservation spans 40–45 amino acids with a highly invariable core of 11 residues at the C-terminal end (Fig 2A and 2B). Since CIC contains several conserved motifs that exert context-dependent functions [37], we tested the requirements of C1 in different *Drosophila* tissues. To this end, we used CRISPR-Cas9 to generate new *cic* alleles specifically disrupting the C1-coding sequence (S1 Fig). Among the isolated mutants, we selected an allele (designated cic^4) that removes four amino acids in the resulting protein, including three highly conserved residues in the C1 core (Fig 2B).

cic^4 homozygous flies are semilethal and show a range of developmental defects. During early embryogenesis, maternal CIC protein normally establishes the presumptive trunk and abdominal regions of the embryo by restricting *tailless* (*tll*) and *huckebein* (*hkb*) expression to the embryonic poles. At the poles, CIC is downregulated by Torso RTK signaling, thereby enabling localized induction of *tll* and *hkb* by broadly distributed activators [5,9,11,12] (reviewed in 4). In *cic* mutant embryos, *tll* expands towards the center of the embryo, which then causes repression of central patterning genes such as *knirps* (*kni*) and loss of central body regions. Consistent with a loss of maternal CIC function, cic^4 females are fully sterile and lay embryos that lack all central thoracic and abdominal segmented regions (Fig 2C and 2D). Indeed, such embryos show clear derepression of *tll* and loss of the central *kni* stripe at the blastoderm stage (Fig 2E and 2F), indicating a failure of CIC-mediated repression. This effect is not caused by reduced CIC protein expression or stability, since cic^4 embryos exhibit normal levels of CIC accumulation in blastoderm nuclei (Fig 2G and 2H), implying that the CIC^4 mutant is functionally defective. A comparison with other *cic* mutations indicates that cic^4 represents a strong hypomorphic allele (S2 Fig).

Next, we assayed the effects of cic^4 in the follicular epithelium of the ovary. In this tissue, CIC organizes the future dorsoventral (DV) axis of the embryo by repressing *mirror* (*mirr*), thereby restricting its expression to dorsal positions. In *cic* mutant backgrounds, *mirr* becomes derepressed towards ventral regions and this leads to inappropriate repression of *pipe*, a gene that is critical for induction of ventral embryonic fates [6,8,38–40]. Consequently, the resulting progeny show a strongly dorsalized phenotype and loss of ventral patterning markers. We find that ovaries from cic^4 females show derepressed *mirr* expression that is similar to that seen in strong *cic* mutant conditions (Fig 2I and 2J) [6,8]. In addition, embryos laid by cic^4 females lack expression of *twist* (*twi*), a target of the maternal DV system that is normally activated in ventral positions (Fig 2K and 2L). Thus, C1 is also required for CIC repressor activity in the follicle cells.

Also, cic^4 flies consistently show abnormal wings with extra vein tissue (Fig 2M and 2N). This phenotype reflects insufficient CIC activity during wing development, where CIC represses vein-promoting genes downstream of EGFR signaling [7,12] (see also below).

PLOS | GENETICS

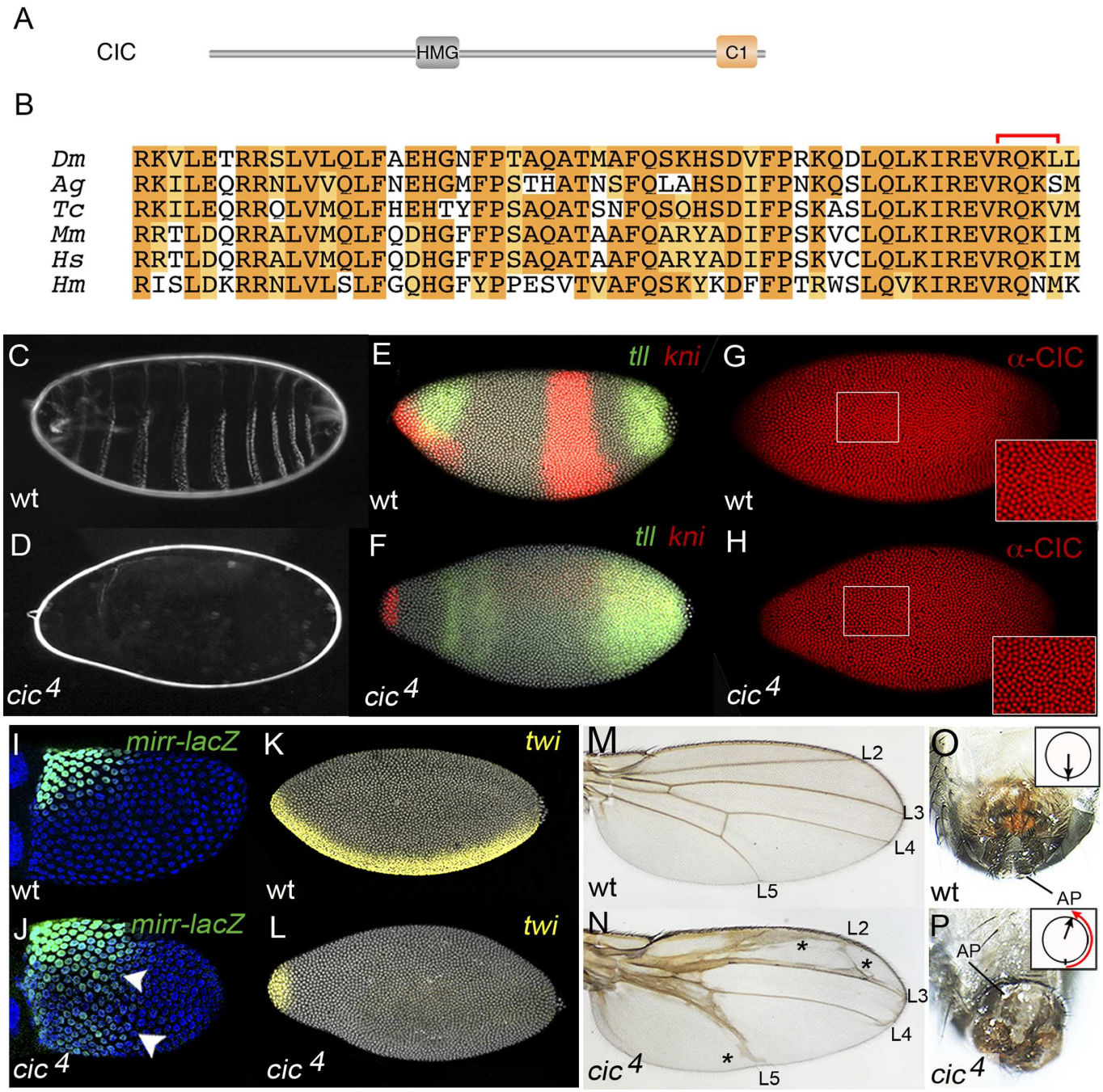


Fig 2. The C1 domain is required for multiple CIC functions in *Drosophila*. (A) Diagram of *Drosophila* CIC protein indicating the positions of the HMG-box and C1 domains. (B) Alignment of C1 domain sequences from *Drosophila* (*Dm*), *Anopheles* (*Ag*), *Tribolium* (*Tc*), mouse (*Mm*), human (*Hs*) and hydra (*Hm*). Light shading indicates similar residues. The four residues deleted by the *cic*^{4} mutation are indicated by a red bracket (see also S1 Fig.). (C, D) Cuticle of embryos derived from wild-type (C) and homozygous *cic*^{4} (D) females. The lack of patterning elements in the mutant reflects both suppression of trunk and abdominal regions as well as complete dorsalization of the embryo (see panels F and L). (E, F) Patterns of *tll* (green) and *kni* (red) mRNA expression in wild-type (E) and *cic*^{4} (F) embryos; nuclei are labeled with DAPI (grey). The mutant embryo shows expanded *tll* expression, which then causes repression of the abdominal *kni* domain. (G, H) Immunodetection of CIC protein in embryos from wild-type (G) and *cic*^{4} (H) females using anti-CIC antibody. Both backgrounds show similar levels of CIC nuclear accumulation (insets), indicating that the CIC^4 mutant is stable but functionally inactive. (I, J) Patterns of *mirr-lacZ* reporter expression (green) in wild-type (I) and *cic*^{4} (J) stage-10 egg chambers; nuclei are labeled with DAPI (blue). Note the ventrally expanded expression of *mirr-lacZ* (arrowheads). (K, L) Expression of *twi* mRNA (yellow) in wild-type (K) and *cic*^{4} (L) embryos; nuclei are labeled with DAPI (grey). *twi* expression is severely reduced in the mutant embryo. Panels E, G, I, J and K are oriented with anterior to

the left, dorsal up. (M, N) Wings from wild-type (M) and *cic*^{4} (N) adult flies; veins L2-L5 are indicated. The mutant displays thickened veins and ectopic vein material (asterisks). (O, P) External genitalia from wild-type (O) and *cic*^{4} (P) males; AP, anal plates. The *cic*^{4} individual exhibits a genital rotation phenotype (arrows indicate the genital arch-to-AP orientation).

doi:10.1371/journal.pgen.1006622.g002

Finally, *cic*^{4} males are sterile and exhibit a severe genitalia rotation phenotype present in other strong *cic* mutants backgrounds [7] (Fig 2O and 2P). Thus, although we have not examined the requirement of C1 for all CIC functions in *Drosophila*, our results suggest that C1 mediates a key general aspect of CIC activity in this organism.

Replacing the HMG-box of CIC by a heterologous DNA binding domain renders C1 dispensable for repression

Since C1 is important for CIC repressor function, we initially hypothesized that C1 might function as a repressor module that interacts with co-repressor factors. As a first test of this idea, we reasoned that replacing the HMG-box region of CIC with a heterologous DNA binding domain should produce a chimeric protein capable of repressing transcription in a C1-dependent manner. For this, we adopted an assay involving the basic-helix-loop-helix (bHLH) region of Hairy and the *Sex-lethal* (*Sxl*) gene [41,42]. We made a *cic* expression construct in which the HMG-box-coding region was replaced by the bHLH-coding region of Hairy and expressed this chimera, CIC(bHLH) (Fig 3A), in transgenic embryos under the control of *cic* genomic sequences. The CIC(bHLH) product was clearly detectable in nuclei of central and subterminal regions of the embryo (Fig 3B), whereas a CIC derivative lacking the HMG-box is mainly cytoplasmic [9], implying that the bHLH region targets CIC to the

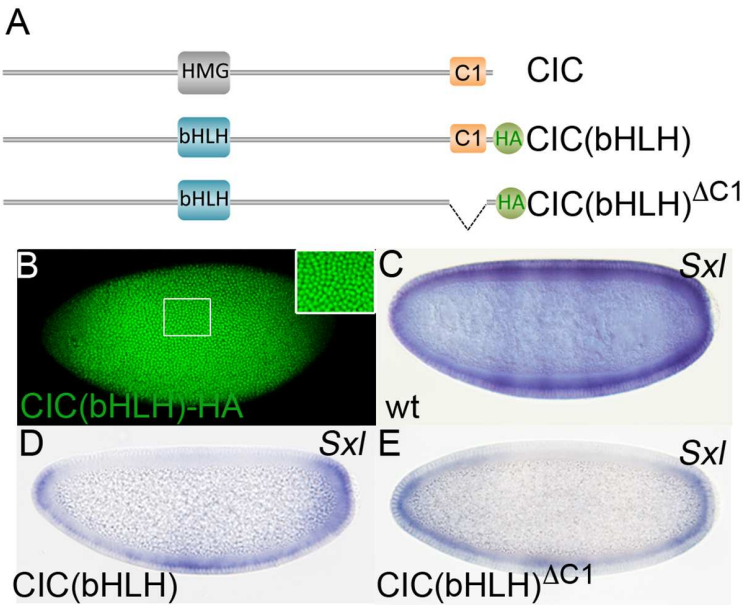


Fig 3. Fusion of CIC to a heterologous DNA binding domain bypasses the requirement for C1. (A) Structure of *Drosophila* CIC and CIC derivatives in which the HMG-box has been replaced by the bHLH domain of Hairy. The CIC(bHLH) and CIC(bHLH)$^{\Delta C1}$ proteins are tagged with the HA epitope and are thus discernable from endogenous CIC. (B) Expression of CIC(bHLH)-HA in embryos stained with anti-HA antibody. The inset shows a higher magnification view of nuclear CIC(bHLH)-HA accumulation. (C-E) *Sxl* mRNA expression in female wild-type (C) and transgenic embryos expressing CIC(bHLH) (D) and CIC (bHLH)$^{\Delta C1}$ (E). *Sxl* appears clearly repressed in both transgenic embryos.

doi:10.1371/journal.pgen.1006622.g003

nucleus. One target recognized by the bHLH domain of Hairy is *Sxl*, a sex-determining gene activated exclusively in female embryos (Fig 3C). Hairy does not normally regulate *Sxl*, but it can do so when expressed earlier than usual (i.e. before or at early stage 5), by mimicking the activity of the Hairy-family repressor Deadpan [41,43]. Thus, premature Hairy expression causes inappropriate repression of *Sxl* and leads to female lethality. Accordingly, we find that early CIC(bHLH) expression driven by the maternal *cic* promoter causes extensive repression of *Sxl* except in polar regions of the embryo, where CIC(bHLH) nuclear levels and activity are lower in response to Torso signaling (Fig 3D). Consistent with this repression effect, CIC (bHLH)-expressing females show a clear 'daugtherless' phenotype as >95% of their progeny are males.

We then tested a CIC(bHLH) derivative lacking the C1 domain (Fig 3A, Materials and methods). Surprisingly, this construct behaves similarly to intact CIC(bHLH), causing evident repression of *Sxl* and lethality of the female progeny (Fig 3E). Thus, targeting CIC to the *Sxl* regulatory sequences via the Hairy bHLH domain renders CIC-mediated repression independent of C1. In concordance, we recently found that fusing C1 directly to a bHLH-containing fragment of Hairy does not lead to repression in the *Sxl* assay [37], whereas similar fusions with well-characterized repressor domains do [42,44–46]. Thus, although we cannot rule out that C1 could have an intrinsic repressor activity in other contexts, the fact that C1 is required for CIC but not CIC(bHLH) function points to a role of C1 in HMG-box-mediated DNA binding.

The HMG-box and C1 domains are both needed for transcriptional repression and promoter targeting in human cells

Next, we tested the function of C1 in human cultured cells. To this end, we made a series of GFP-tagged human CIC derivatives carrying mutations in the HMG-box and C1 domains (Fig 4A). These constructs had similar levels of expression and were all detected in the nucleus, implying that tagging and mutagenesis did not differentially affect their stability or subcellular localization (Fig 4B; S3 Fig). Then, we assayed the repressor activities of the different constructs using a luciferase reporter under the control of a synthetic promoter carrying CIC binding sites (CBSs) derived from the *ETV5* promoter [15,17,31] (see Fig 4 legend). This reporter, *ETV5p*, was significantly repressed upon cotransfection of GFP-CIC (WT), whereas GFP-CIC constructs carrying either recurrent OD mutations affecting the HMG-box (R215W) or C1 (R1515L) domains [22,26,30], or a complete C1 deletion (ΔC1), showed reduced repressor activities relative to the intact control (Fig 4C). This indicates that the HMG-box and C1 domains are both required for CIC repressor activity in mammalian cells.

We then tested association of the above CIC mutants to endogenous *ETV/PEA3* gene promoters by chromatin immunoprecipitation (ChIP). As expected, the R215W mutation caused a reduction in ChIP signals relative to the control CIC protein (Fig 4D). Notably, both C1 mutations also diminished CIC promoter occupancy at all three *ETV/PEA3* gene analyzed. The reduction was more pronounced for the full C1 deletion, but the effect of the R1515L mutation was also clearly significant. These results indicate a requirement of both the HMG-box and C1 domains for binding of CIC to its target genes.

The C1 domain cooperates with the HMG-box in DNA binding

Having established that C1 mediates CIC association with endogenous targets, we hypothesized that it might contribute directly to DNA binding. To test this idea, we performed EMSA experiments comparing the binding activities of various *Drosophila* and human CIC constructs. As shown in Fig 5A, these constructs carried intact or mutated HMG-box and C1

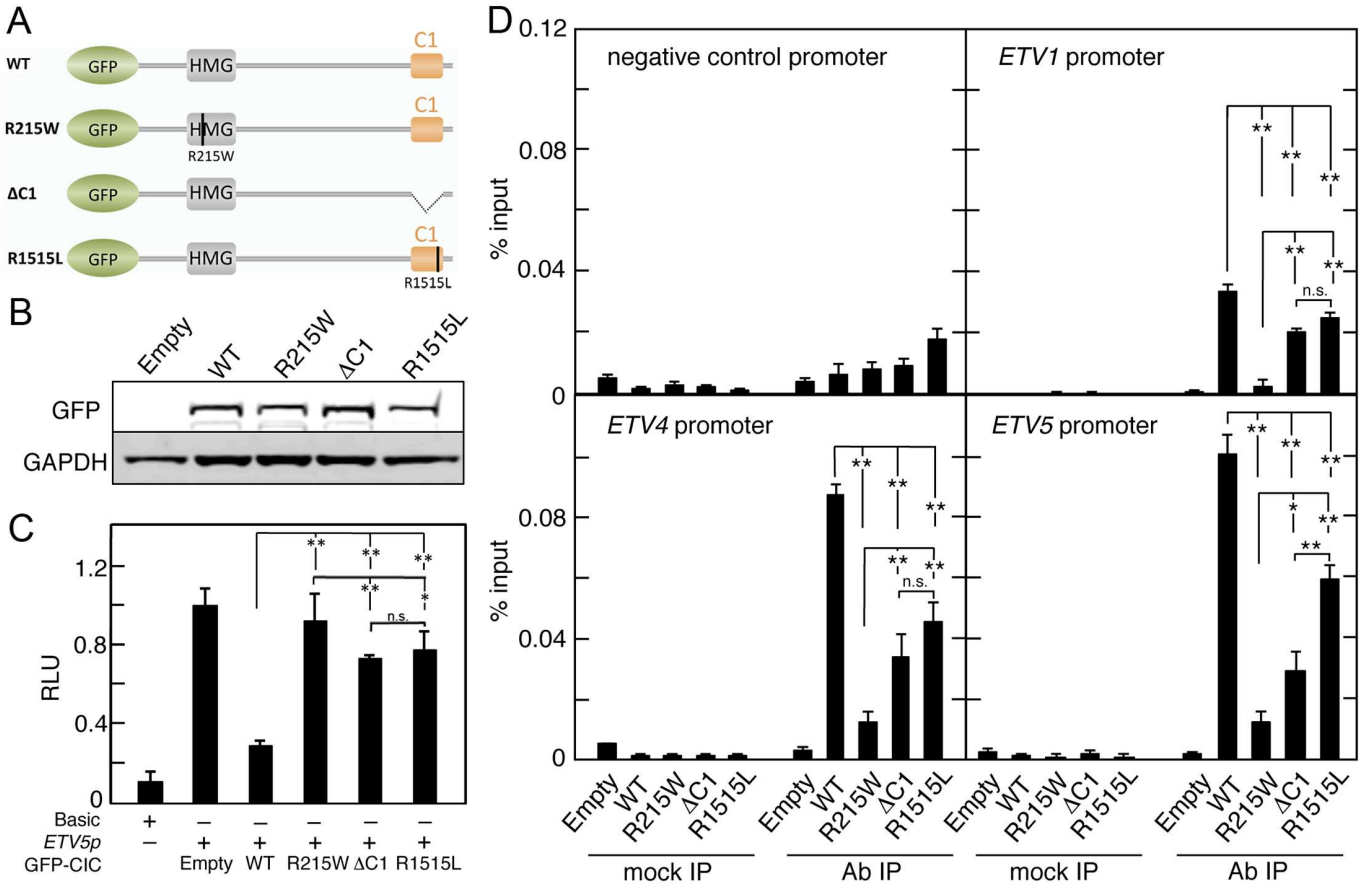


Fig 4. The C1 domain mediates CIC repression and promoter binding in human cells. (A) Diagram of GFP-tagged human CIC protein constructs tested in reporter and ChIP assays. Mutations in the HMG-box and C1 domains are indicated by vertical lines in both domains. (B) Western blot analysis of wild-type and mutant GFP-CIC fusion proteins stably expressed in Flp-In T-REx 293 cells using antibodies directed against GFP. GAPDH expression served as a loading control. (C) Relative luciferase expression levels driven by a promoter-less vector (*Basic*) or a synthetic promoter carrying CIC binding sites derived from the *ERM/ETV5* promoter (*ETV5p*), in the absence or presence of wild-type (WT) or the indicated mutant GFP-Cic constructs transfected into 293T cells. Luciferase values are expressed relative to the activity of the reporter co-transfected with empty *pcDNA5/FRT/TO* vector (Materials and methods). (D) ChIP assay using GFP antibodies in Flp-In T-REx 293 cells stably expressing wild-type (WT) or the indicated mutant GFP-CIC fusion proteins. Flp-In T-REx 293 cells stably transfected with an empty vector were used as a control (Empty). Association with the CIC binding elements in the *ETV1*, *ETV4* and *ETV5* promoters was analyzed by quantitative real-time PCR and normalized to the amount of input DNA. Statistical analysis was performed with one-way ANOVA followed by Tukey's *post hoc* test; (*P<0.05 and **P<0.01); n.s., non significant.

doi:10.1371/journal.pgen.1006622.g004

domains either alone or combined together in the same polypeptide. These proteins were expressed in vitro or in bacteria, and incubated with probes derived from *Drosophila* and human CIC targets. Unexpectedly, the *Drosophila* HMG-box alone (construct 1) was unable to bind DNA, whereas an HMG-C1 construct containing both domains next to each other (construct 2) showed clear, specific binding to a probe from the *hkb* gene containing two CBSs [12]. Similarly, neither the HMG-box nor the C1 domains alone (constructs 1 and 3) bound to a probe from *pointed* (*pnt*) [14], nor did they bind this probe when combined in the same reaction (Fig 5A and 5B). In contrast, this probe was readily bound by a His-tagged HMG-C1 construct purified from bacteria, but not by the equivalent construct bearing the cic^4 lesion (constructs 4 and 5). Likewise, a human HMG-C1 construct efficiently bound to a probe from the *ETV4* gene [15,17], whereas recurrent OD mutations mapping to the HMG-box (R201W and R215W) or C1 (R1515L) domains greatly reduced this binding (constructs 6–9). We also

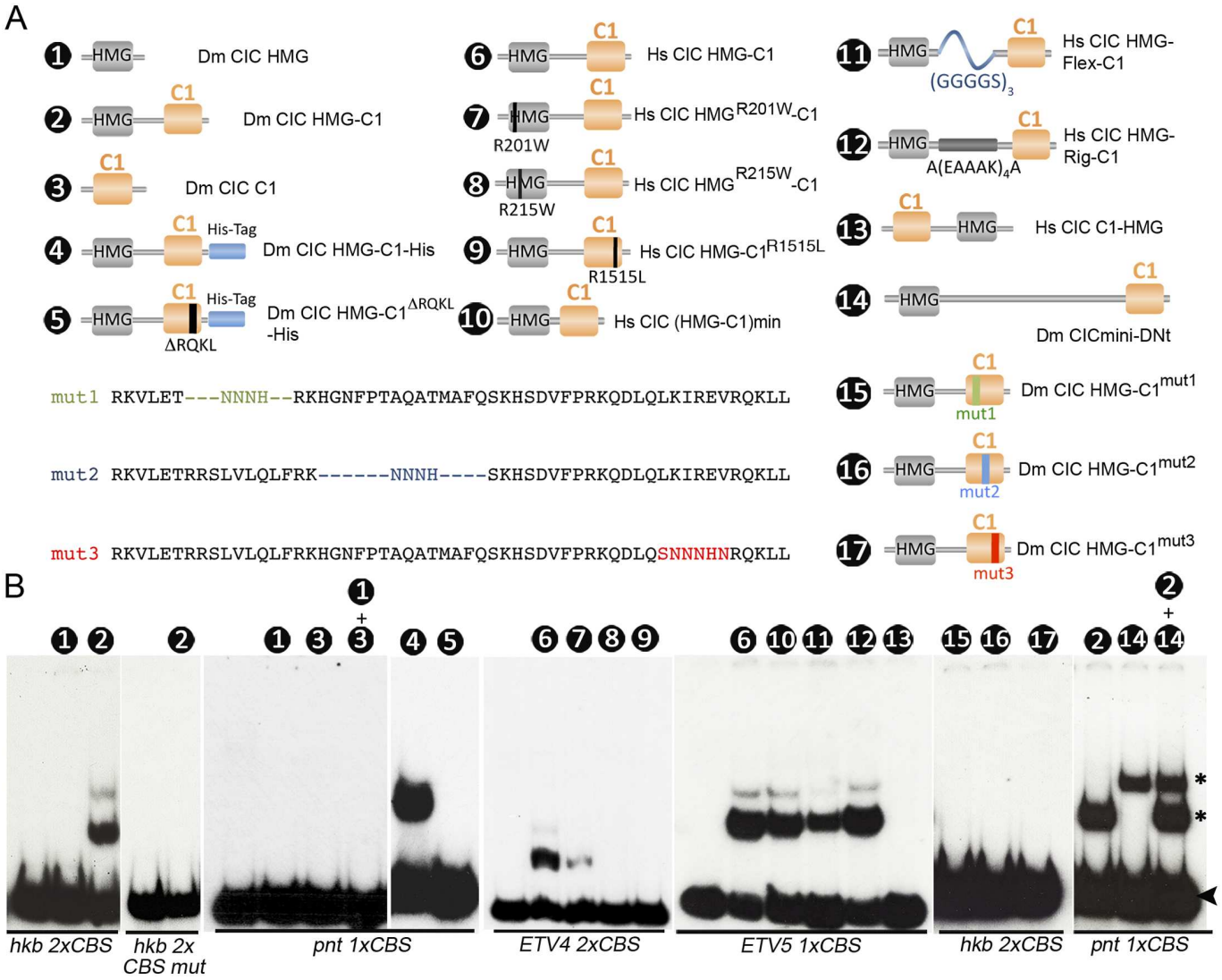


Fig 5. The HMG-box and C1 domains are both essential for binding of CIC to DNA. (A) Diagram of CIC protein constructs tested in EMSA experiments. Constructs 1–3 and 6–17 were transcribed and translated in vitro; constructs 4 and 5 were expressed and purified from bacteria. Construct 2 contains the HMG-box and C1 domains in close proximity, without the intervening sequences that normally separate both domains (see S4 Fig showing that this arrangement is functional in vivo). Construct 10 represents a minimal (min) version of construct 6 where the HMG-box and C1 domains have been placed immediately next to each other. Dashes in the partial sequences of constructs 15 and 16 indicate deleted residues. (B) EMSA analyses of CIC constructs binding to different wild-type and mutant DNA probes. Numbers indicate the constructs used in the binding reactions; unlabeled lanes contain unprogrammed reticulocyte lysate as a negative control. The probes used are indicated below the gels; *1xCBS* and *2xCBS* indicate the presence of 1 or 2 endogenous CIC octameric sites, respectively. *hkb 2xCBS mut* carries mutated CIC sites. The arrowhead marks the position of free, unbound probe in all the gels. Asterisks indicate the differential mobility of protein:DNA complexes. The sequences of wild-type and mutant probes are shown in S1 Table.

doi:10.1371/journal.pgen.1006622.g005

used human HMG-C1 constructs to test the effects of flexible versus rigid linkers separating the HMG-box and C1 domains (constructs 10–12; see Fig 5 legend); both linkers permitted effective binding, ruling out a major effect of the sequences connecting the HMG-box and C1 elements. In contrast, placing the HMG-box and C1 domains in reverse order (construct 13) abolished DNA binding (Fig 5B), indicating that this configuration imposes steric constraints on binding. Thus, the HMG-box and C1 domains function together as an obligate, conformationally oriented module for site-specific binding to DNA.

As indicated above, the C1 sequence contains a highly conserved core of 11 residues flanked by a somewhat more variable amino-terminal extension. To further test the requirements of these sequences in DNA binding, we assayed three mutations of discrete sub-motifs within C1 (constructs 15–17). All three mutations prevented binding of *Drosophila* HMG-C1 to the *hkb* probe (Fig 5B), indicating that these sub-motifs (or a full, correctly folded C1 domain) are important for function.

Finally, we asked if, by analogy to other Sox factors [47–49], the HMG-C1 module binds DNA as a homodimer. To this end, we assayed the binding activities of two HMG-C1 constructs of different size (constructs 2 and 14) using the *pnt* probe, which contains a single CBS. As expected from their relative molecular masses (approximately 24 and 53 kD, respectively), each of these constructs individually produced protein-DNA complexes of different mobility. Similarly, a binding reaction containing both proteins resulted in the same complexes and no intermediate complex was observed (Fig 5B), indicating that the proteins did not oligomerize. These results strongly suggest that C1 does not mediate dimerization and the HMG-C1 module binds DNA as a monomer, although we cannot formally exclude that other CIC sequences may facilitate oligomerization during DNA binding in vivo.

The HMG-C1 module recognizes discrete octameric sites during DNA binding

The role of C1 in DNA binding is reminiscent of the mechanism employed by certain TCF factors in DNA recognition. Thus, the TCF orthologs from *Drosophila* and *C. elegans*, and some vertebrate TCF isoforms, contain, in addition to the HMG-box, a zinc binding domain known as C-clamp which functions in DNA binding. The C-clamp acts by binding so-called 'Helper sites' (5'-RCCGCCR-3') located at short distance (usually <10 bp away) from the sequence recognized by the TCF HMG-box, thereby augmenting the DNA binding strength and specificity of TCF towards its targets [50–56]. Therefore, although the C1 and C-clamp domains are not related in sequence, we considered the possibility that C1 might also recognize a specific conserved motif adjacent to the consensus CBSs. To this end, we first compared the sequences flanking bona fide CBSs present in three *D. melanogaster* genes, their *D. virilis* orthologs, and three mouse promoters. As shown in Fig 6A, this analysis reveals several conserved motifs in the vicinity of CIC octamers from orthologous *Drosophila* genes, but not across non-orthologous genes. This suggests that those motifs correspond to orthologous sites for other transcription factors in the selected enhancers or promoters. Similarly, the mouse CBSs are flanked by a short A/T-rich extension, but this motif is not well conserved in the *Drosophila* sequences. This indicates that CIC sites are not surrounded by a particular motif serving as a 'helper site' for CIC DNA binding.

To directly test the influence in DNA binding of sequences flanking functional CIC octamers, we performed EMSA experiments using probes corresponding to CIC sites present in the *Drosophila tll*, *hkb* and *intermediate neuroblasts defective* (*ind*) genes, and in human *ETV5*. These probes span 30–32 bp and do not share significant similarity outside the CIC octamers (Fig 6B). Nevertheless, they were similarly bound by the corresponding *Drosophila* and human HMG-C1 minimal proteins, indicating that the CIC octamer is the main determinant for DNA recognition in this assay (Fig 6C and 6D). The human protein also bound efficiently a synthetic probe containing a CIC octamer flanked by random sequences (*CBS syn*). Finally, we tested the binding of human HMG-C1 to an 18-bp probe carrying a CBS derived from *ETV5* flanked by only 5 bp on either side (Fig 6B). This probe was bound with similar affinity to that observed using longer probes, and the binding was reduced by a

PLOS | GENETICS

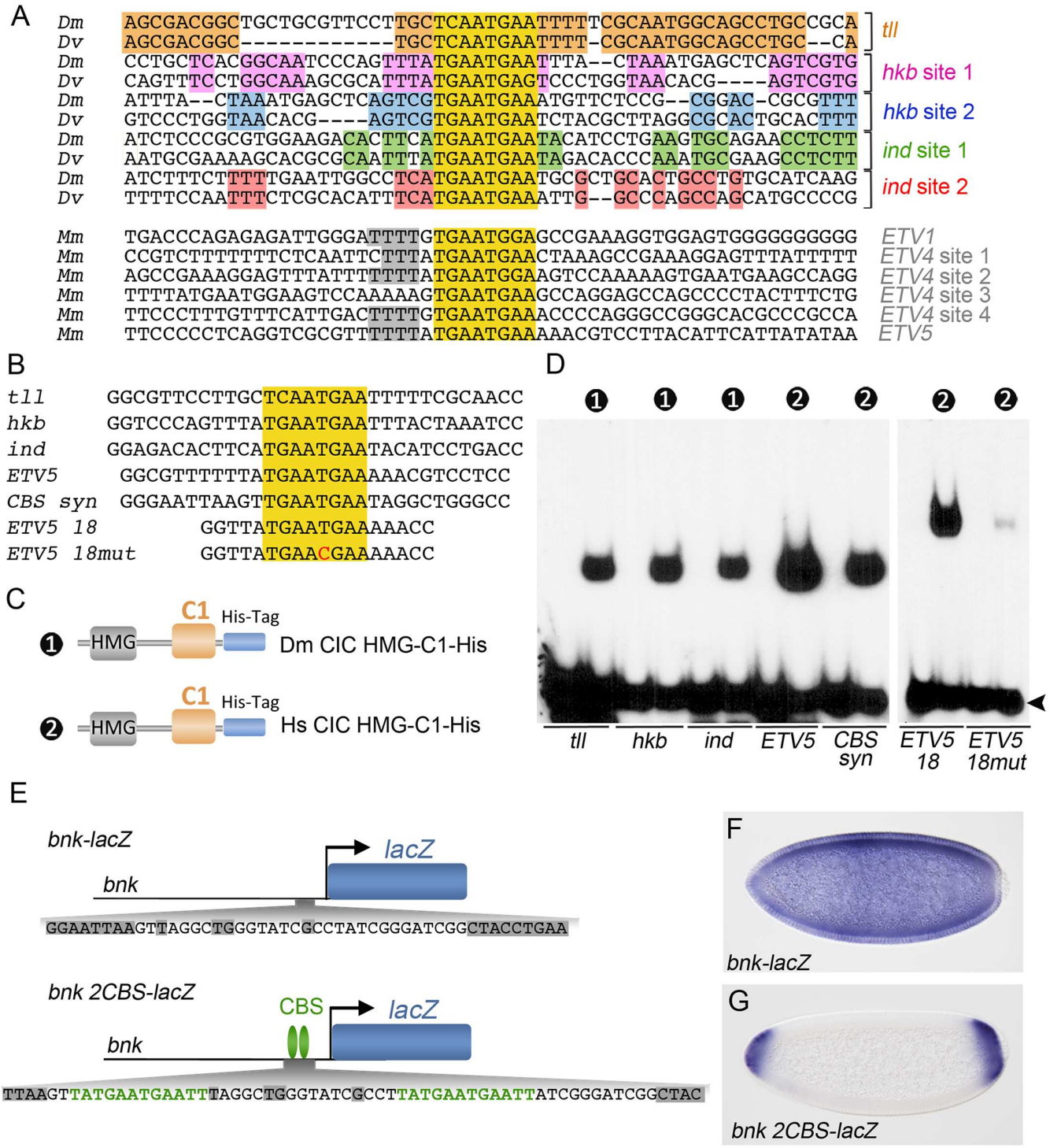


Fig 6. CIC recognizes individual octameric sites and does not depend on helper sites for selecting its targets. (A) Alignment of sequences flanking functional CBSs from selected *D. melanogaster* (*Dm*), *D. virilis* (*Dv*) and mouse (*Mm*) CIC target genes. The CBSs are highlighted in yellow. Conserved flanking motifs are shaded in different colors. (B) Sequences of probes containing intact or mutated CBSs. (C) Diagram of recombinant *Drosophila* (Dm) and human (Hs) CIC constructs used in the EMSA experiments; both constructs were produced in bacteria. (D) EMSA analyses using the DNA probes and proteins shown in panels B and C, respectively. Numbers indicate the constructs used in the binding reactions; unlabeled lanes are negative controls without added protein. Free probes are indicated by an arrowhead. (E) Diagram of a control *bnk-lacZ* reporter and a modified

version carrying two CBSs (*bnk 2CBS-lacZ*). The positions of the inserted CBSs are indicated below the reporters, with conserved motifs among *Drosophila* species shaded in grey. (F, G) Patterns of expression of *bnk-lacZ* and *bnk 2CBS-lacZ* reporters.

doi:10.1371/journal.pgen.1006622.g006

mutation in the CBS (Fig 6D), indicating that a single, isolated CIC octamer is sufficient for effective binding of CIC to DNA.

Finally, we have re-examined if CBSs are sufficient for DNA recognition by CIC in vivo. We selected a 12-bp motif containing a CBS from the *hkb* enhancer and inserted two copies of this sequence in a reporter construct driven by the *bottleneck* (*bnk*) promoter, which is ubiquitously active in early embryos. These insertions were introduced without disrupting conserved elements in the *bnk* promoter, thus preserving its regulation by the Zelda activator and other factors (Fig 6E) [57]. As shown in Fig 6F and 6G, whereas a control *bnk-lacZ* reporter directs uniform expression in the early embryo, the reporter containing CBSs is expressed only in polar regions, indicating that it is effectively regulated by endogenous CIC. This result supports our conclusion that CIC binds its target sites without any requirement or modulation by specific flanking sequences.

C1-dependent activity of a CIC-DUX4 chimera in *Drosophila*

The above results provide a plausible mechanistic explanation for the main pattern of oncogenic CIC-DUX4 chimeras (which usually include the C1 domain, as shown in Fig 1), since C1 should promote CIC-DUX4 activity by enhancing its binding to DNA. Pursuing this idea, we have established a *Drosophila* assay of CIC-DUX4 activity in which to test the requirement of C1. We made a construct encoding *Drosophila* CIC fused to the C-terminal portion of human DUX4 and expressed this chimera in the developing wing (Fig 7A). This tissue is highly sensitive to changes in CIC activity, which normally acts to promote intervein cell fate except in the presumptive veins where it is inhibited by EGFR signaling (Fig 7B). Thus, loss of CIC function produces extra vein material (Fig 2N; S2 Fig), whereas overexpression of CIC suppresses vein formation (Fig 7D).

We find that targeted expression of CIC-DUX4 in the primordial wing blade (see Materials and methods) causes severe defects including reduced wing size, ectopic venation and blistered wings due to loss of adhesion between the two wing surfaces (Fig 7E and 7F). This phenotype is markedly different to that caused by overexpression of intact CIC and actually resembles the loss of CIC function (see S2 Fig), consistent with CIC-DUX4 mediating transcriptional activation instead of repression. To test this further, we assessed CIC-DUX4 activity in the wing imaginal disc using a synthetic reporter, *CUASC-lacZ*, containing CBSs linked to GAL4 binding sites (Fig 7I). In discs expressing GAL4 protein in the wing pouch, the reporter is activated only in presumptive vein stripes since it is repressed by CIC in intervein regions (Fig 7J) [12]. In contrast, this pattern appears markedly broadened in CIC-DUX4-expressing discs (Fig 7K), as expected if CIC-DUX4 activates the reporter and overrides the repressor activity of endogenous CIC. We then evaluated the contribution of the C1 domain to CIC-DUX4 activity in this assay. Using CRISPR-Cas9, we edited the CIC-DUX4-expressing transgene and isolated two mutations deleting either 2 or 11 residues within the C1 domain of CIC-DUX4 (Fig 7A). Both mutations strongly suppressed the phenotypes produced by CIC-DUX4, with the 11-residue deletion showing almost complete restoration of the wild-type vein pattern (Fig 7G and 7H). This mutant also showed significantly restricted expression of the *CUASC-lacZ* reporter (Fig 7M). Thus, the C1 domain is required for the opposing activities of CIC and CIC-DUX4 proteins in the *Drosophila* wing, which is consistent with its role in DNA binding rather than transcriptional repression per se.

PLOS | GENETICS

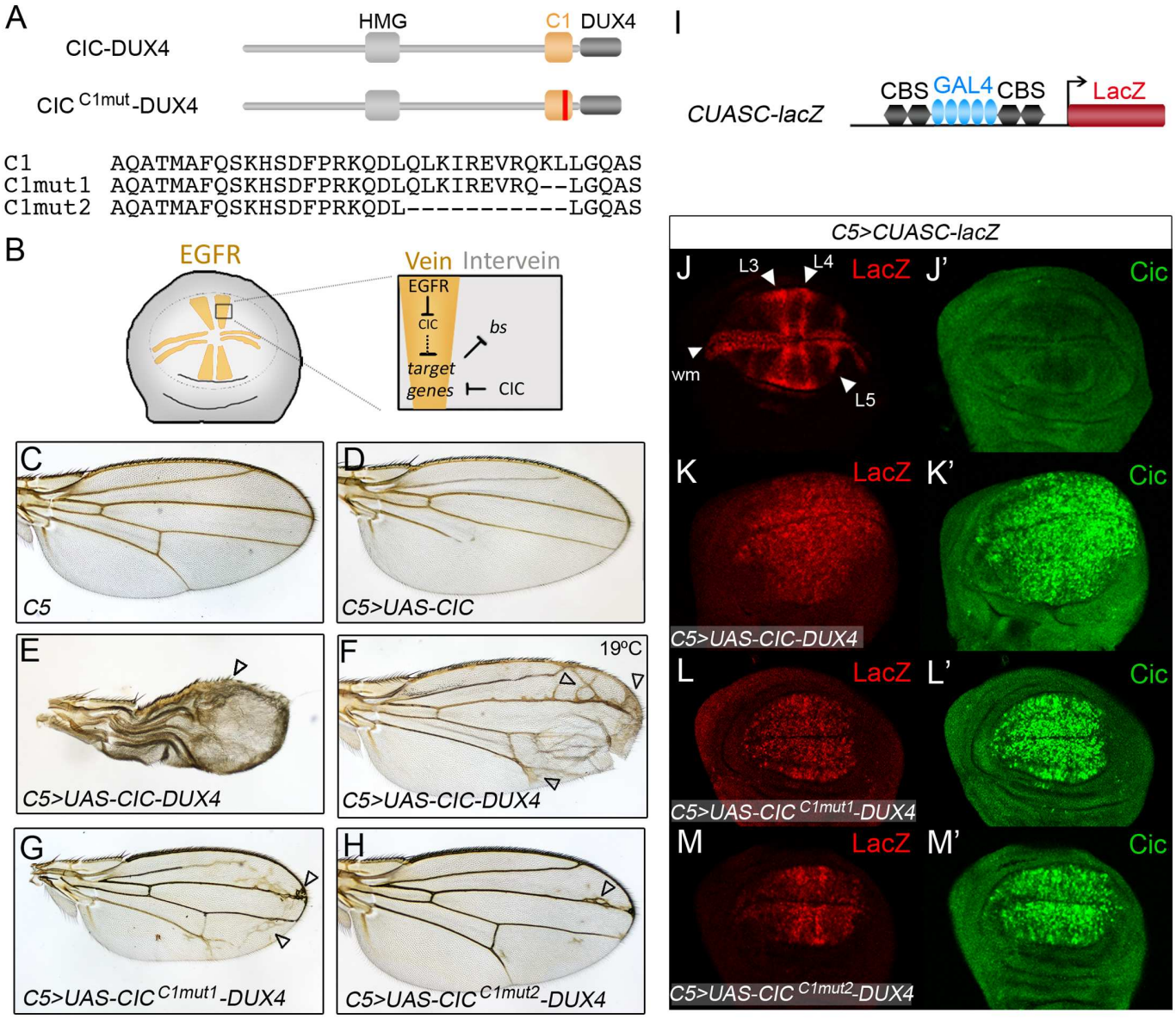


Fig 7. The C1 domain is required for the activity of a CIC-DUX4 fusion in the *Drosophila* wing. (A) Diagram of CIC-DUX4 chimeras expressed in the wing using the GAL4-UAS system. CIC^{C1mut}-DUX4 indicates two different derivatives carrying CRISPR-Cas9-induced mutations (vertical red line) in the C1 domain; partial sequences of intact and mutant C1 domains are shown below. (B) Model of CIC function in the wing primordium. The pattern of wing veins is established in the imaginal disc through localized activation of the EGFR signaling pathway, which downregulates CIC in presumptive vein cells (yellow). CIC in turn promotes the intervein fate by repressing (directly or indirectly) EGFR-induced genes such as *ventral veinless* and *decapentaplegic*, while indirectly maintaining *blistered* (*bs*) expression in intervein cells [7,12,67]. (C-H) Wing phenotypes induced by expression of CIC (D), CIC-DUX4 (E, F) and two CIC-DUX4 mutant derivatives carrying deletions in the C1 domain (G, H) under the control of the *C5* GAL4 driver; a control wing with GAL4 driver only is shown in C. Arrowheads indicate broadened veins and ectopic vein material in CIC-DUX4-expressing wings. Unless otherwise indicated, all panels were obtained by raising flies at 25˚C; panel F shows the weaker phenotype resulting from induction of CIC-DUX4 at 19˚C. (I) Diagram of the *CUASC-lacZ* reporter driven by a synthetic enhancer composed of five GAL4 binding sites flanked by two CBSs on either side. (J-M') Late third-instar wing discs doubly stained with anti-lacZ (J-M) and anti-Cic antibodies (J'-M'). J and J' show a wild-type disc carrying the *CUASC-lacZ* reporter. K-M' show representative discs expressing intact and mutant CIC-DUX4 proteins using the *C5* driver, which is expressed in the wing pouch and serves to activate both the effector genes and the *CUASC-lacZ* reporter.

doi:10.1371/journal.pgen.1006622.g007

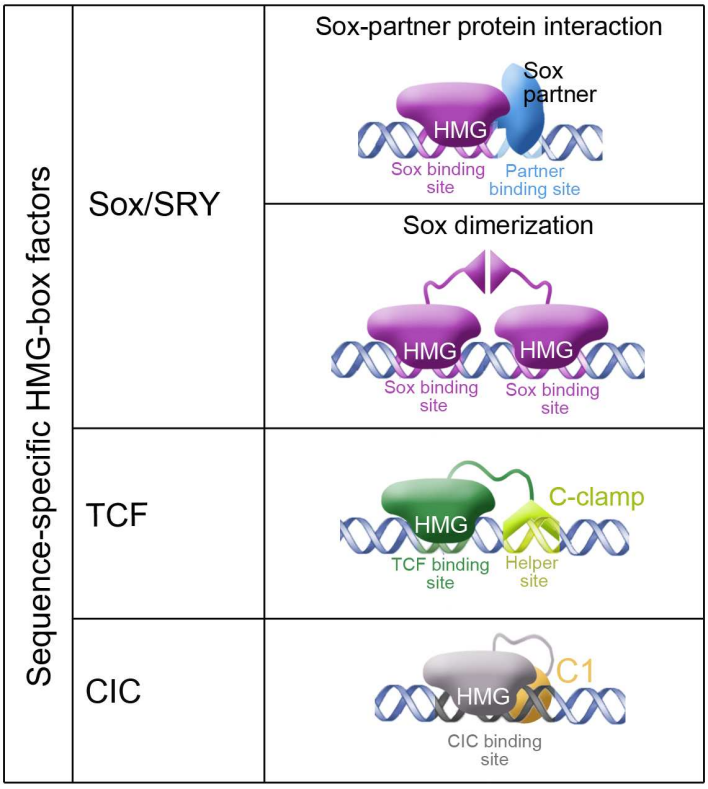


Fig 8. Distinct modes of target recognition by sequence-specific HMG-box proteins. The diagram summarizes the main DNA-binding mechanisms used by each HMG-box sub-family. Sox proteins usually bind their Sox sites in combination with partner factors that recognize adjacent DNA sequences, but can also form homo- and heterodimers via specific dimerization motifs such as those present in SoxD and SoxE family members. Some TCF factors also exhibit bi-partite DNA recognition via the HMG-box and the C-clamp domain that binds GC-rich sequences known as Helper sites. In contrast, CIC proteins appear to bind individual octameric sites through their HMG-box and C1 domains, acting independently of other specific DNA sites and partner proteins.

doi:10.1371/journal.pgen.1006622.g008

Discussion

HMG-box proteins play critical roles in development and disease by regulating the expression of specific target genes. For both Sox and TCF factors, this control depends on the HMG-box as well as on other DNA-binding and dimerization motifs that cooperate in regulating the correct genomic targets. For instance, Sox proteins typically associate with partner factors that interact with specific DNA sequences close to the Sox sites. Similarly, several TCF isoforms contain a C-clamp domain that recognizes GC-rich motifs adjacent to TCF sites, thereby enhancing the affinity and specificity of TCF binding to its targets. It is believed that such combinatorial modes of DNA recognition are essential for proper developmental regulation by both protein families (see Fig 8).

In this work, we have identified a distinct mode of DNA binding by CIC, which depends on its conserved C1 domain. Compared to the above examples, the C1 motif is unique in that it is located at long distance from the HMG-box, does not display detectable DNA-binding activity on its own, does not mediate dimerization, and is not involved in recognizing auxiliary motifs next to CIC octameric sites. Instead, our results indicate that C1 cooperates with the HMG-box to recognize discrete octameric sites both in vitro and in vivo. Since mutations in the C1

domain do not completely abolish the activity of CIC in flies or in human cells (S2 Fig; Fig 4), we favor the view that C1 acts by potentiating the binding of the HMG-box to its specific sites. Several mechanistic models could account for C1 function. For example, C1, like many DNA binding sequences, contains several conserved basic residues in its core, which might establish direct, low-affinity contacts with DNA. Alternatively, the C1 domain could interact with the HMG-box or modulate its folding during DNA recognition. Future high-resolution structural analyses of the HMG-box-C1 module bound to DNA should elucidate the molecular basis of C1 function.

Regardless of the precise molecular mechanism, our results reveal a unique mode of DNA binding that distinguishes CIC from Sox and TCF factors (Fig 8). Thus, the HMG-box-C1 module mediates robust and specific binding to its conserved octameric sites independently of partner factors and auxiliary target sequences. Indeed, CIC recognizes their octameric sites even when those sites are relocated to heterologous or synthetic enhancers (Fig 6G; see also refs. [11,12,19,31]), and our current work demonstrates efficient binding of the HMG-box-C1 polypeptide to an isolated CIC octamer in vitro. It thus appears that HMG-box proteins share a general principle of augmenting their target specificity through modular or cooperative DNA binding, but each individual HMG-box family relies on unique domains and mechanisms for this activity. Furthermore, the distinct binding modes of Sox and CIC proteins give rise to different logics of transcriptional control. Thus, the 'partner mechanism' of Sox proteins is highly versatile and leads to either transcriptional activation or repression depending on the partner protein as well as on the promoter context. In contrast, in all cases studied so far, CIC proteins function as dedicated repressors, and *Drosophila* CIC has been shown to contain an intrinsic repressor motif [37].

Finally, our results imply that the two main subgroups of CIC amino acid substitutions in OD and other tumors, which map to the HMG-box and C1 domains (Fig 1A), cause related defects in DNA binding. This would then lead to derepression of CIC targets such as *ETV/PEA3* genes, which encode ETS transcription factors extensively implicated in tumorigenesis, as well as genes encoding feedback inhibitors of RTK signaling like Sprouty and Spred [23,29,30]. Moreover, our findings help explain the main pattern of oncogenic translocations resulting in CIC-DUX4 sarcomas (Fig 1B): it is not incidental that C1 is preserved in most CIC-DUX4 chimeras, since C1 should be required for effective CIC-DUX4 DNA binding and subsequent aberrant activation of *ETV* genes and other targets. This is supported by our analyses (Fig 7) showing that an intact C1 domain is required for the activity of a CIC-DUX4 chimera in the *Drosophila* wing.

Materials and methods

Drosophila genetics and transgenic lines

The cic^{4} allele was generated by CRISPR-Cas9-mediated editing. Briefly, a custom gRNA expression construct targeting the C1 coding sequence was prepared in vector *pCDF3* [58] and inserted at the *attP40* landing site via phiC31-mediated integration [59] (see S1 Fig for details of the gRNA sequence). Transgenic gRNA males were crossed to *nanos-cas9* females to obtain founder males, which were then crossed to females carrying the *TM3* balancer for recovery of mutant alleles. Induced mutations were characterized by sequencing PCR fragments amplified from candidate flies. A similar scheme using the same gRNA insertion was employed to isolate mutations in the *UAS-CIC-DUX4* transgene. Other alleles and chromosomal rearrangements employed were: cic^{Q474X} [10], cic^{1} [5], *Df(3R)ED6027* (see FlyBase), and the $mirr^{P2}$ enhancer trap (*mirr-lacZ*; ref. [60]). Transgenic flies expressing CIC derivatives were obtained by P-element transformation. Expression of CIC-DUX4 derivatives was achieved using the GAL4-

UAS system and the driver line *C5* [61]. All crosses were performed at 25˚C, unless otherwise noted.

Histochemistry

Embryos were fixed in 4% formaldehyde-PBS-heptane using standard procedures. Ovaries and wing discs were dissected in PBS and fixed with 4% formaldehyde-PBS. In situ hybridizations were performed using digoxigenin-UTP (*kni*, *twi* and *Sxl*) or biotin-UTP (*tll*) labeled antisense RNA probes, followed by incubation with fluorochrome-conjugated anti-digoxigenin or anti-biotin antibodies for FISH analysis, or with secondary antibodies coupled to alkaline phosphatase (AP) for histochemical detection. *Drosophila* CIC was detected using either a guinea pig polyclonal antibody raised against the C-terminal region of the protein [14], or a rabbit polyclonal recognizing the HMG-box and C-terminal regions. Lac-Z and HA-tagged proteins were detected using monoclonal antibodies 40-1a (Developmental Studies Hybridoma Bank) and 12CA5 (Roche), respectively. Immunofluorescence signals were visualized with species-specific secondary antibodies labeled with different fluorochromes (Molecular Probes). Fluorescent and AP-stained samples were mounted in Fluoramount and Permount, respectively. Cuticle preparations were mounted in 1:1 Hoyer's medium/lactic acid and cleared overnight at 60˚C. Wings were rinsed in isopropanol and mounted in Euparal.

Constructs

The reference sequences used for the *Drosophila* and human CIC proteins are NP_524992.1 and NP_055940.3, respectively. The *CIC(bHLH)* and *CIC(bHLH)*$^{\Delta C1}$ constructs were made using a genomic *cic-HA* rescue transgene in the *pCaSpeR4* vector [9,62], by replacing an EagI fragment encoding amino acids 384–583 of CIC (including the HMG-box) with a fragment encoding residues 25–150 of Hairy (containing the bHLH domain). *CIC(bHLH)*$^{\Delta C1}$ carries, in addition, a deletion of the region coding for the C1 domain (residues 1308–1356). The *CIC-DUX4* transgene encodes most of *Drosophila* CIC protein (residues 1–1380) fused to amino acids 325–424 of DUX4 (thus mirroring the chimera described in ref. 31), and was assembled in *pUAST*.

The constructs used in the EMSA experiments express the following CIC amino-acid fragments: 478–572 (Dm CIC HMG), 478–572 fused to 1288–1378 (Dm CIC HMG-C1), 1288–1378 (Dm CIC C1), 188–288 fused to 1451–1527 (Hs CIC HMG-C1), 188–280 fused to 1457–1527 (Hs CIC (HMG-C1)min), 1457–1527 fused to 188–280 (Hs CIC C1-HMG), and 475–598 fused to 1044–1378 (Dm CICmini-DNt). Hs CIC HMG^{R201W}-C1, Hs CIC HMG^{R215W}-C1 and Hs CIC HMG-C1^{R1515L} are mutant derivatives of Hs CIC HMG-C1. Hs CIC HMG-Flex-C1 and Hs CIC HMG-Rig-C1 are identical to Hs CIC (HMG-C1)min except in that they contain flexible (Flex) and rigid (Rig) linkers separating the HMG-box and C1 domains [63,64]. Dm CIC HMG-C1^{mut1-3} are derivatives of Dm CIC HMG-C1. All these constructs were subcloned into *pET-17b* for in vitro expression under the control of the T7 promoter. His-tagged constructs were expressed in bacteria using the *pET-29b* vector. Dm CIC HMG-C1-His and Dm CIC HMG-C1$^{\Delta RQKL}$-His are derivatives of Dm CIC HMG-C1; Hs CIC HMG-C1-His is based on Hs CIC HMG-C1.

GFP-tagged human CIC constructs were assembled in *pcDNA5/FRT/TO* [15]. The R215W and R1515L mutations were introduced using the QuikChange site directed mutagenesis kit (Agilent) following the manufacturer's guidelines. The C1 deletion (spanning residues 1464–1519) was generated using a recombinant PCR-based approach. Unless indicated otherwise, all plasmids were stably introduced into Flp-In T-REx 293 cells (Invitrogen) following instructions from the manufacturer.

Protein analyses and immunostaining of human cells

For Western blot analysis, cells were lysed in a buffer containing 75 mM NaCl, 50 mM Tris-HCl, pH 8, and 0.5% Triton X-100, supplemented with PMSF and protein inhibitor cocktail Complete Mini (Roche). 50 μg of total protein extract was resolved by SDS-PAGE, transferred to nitrocellulose membranes and probed with antibodies against GFP (Abcam, ab290) and GAPDH (Sigma Aldrich, G8795). To analyze nuclear or cytoplasmic localization of the different CIC constructs, we transiently transfected a plasmid encoding GFP (*pEFGP-C2*) as a control or plasmids encoding WT [15] or mutated GFP-CIC constructs into 293T cells. 48h after transfection, cells were fixed with 4% formaldehyde and permeabilized with 0.5% Triton X-100. GFP expression was detected using polyclonal anti-GFP antibodies (Abcam ab290, 1:1000) followed by counterstaining with Hoechst 33342. Images were acquired with a Leica TCS SP5 confocal microscope.

Luciferase assay

Luciferase assays were performed in a Glomax luminometer (Promega) according to the manufacturer's guidelines. Briefly, we transfected the *pGL3proERM-338/-329* tandem reporter vector [31] along with empty *pcDNA5/FRT/TO* vector or *pcDNA5/FRT/TO* plasmids expressing wild-type or mutant GFP-tagged human CIC derivatives into 293T cells using jetPRIME reagent (Polyplus-transfection). Cells were lysed after 48 h and assayed for luciferase activity. A Renilla luciferase-expressing vector was used for normalization.

ChIP assay

ChIP assays were performed as described [65]. Briefly, $2x10^7$ Flp-In T-REx 293 cells stably transfected with *pcDNA5/FRT/TO* alone or *pcDNA5/FRT/TO* expressing either wild-type or mutated (R215W, R1515L or C1 domain deletion) GFP-tagged human CIC cDNAs were cross-linked for 15 min at room temperature. After washing, cells were sonicated at high intensity during 30 cycles, with 30 s ON and 30 s OFF per cycle (Bioruptor Plus, Diagenode), followed by centrifugation for 15 min at 14,000 rpm at 15˚C. For each condition, 200 μg of lysate was incubated overnight with 2 μl of anti-GFP antibody (Abcam, ab290) and immunoprecipitated by incubation with 20 μl of protein A/G beads during 1 h at 4˚C in a rotating platform. After reverse crosslinking, DNA fragments were recovered by phenol/chloroform extraction and qRT-PCR was carried out in a 7500 Fast Real-Time PCR System (Applied Biosystems) using Power SYBR green PCR Mastermix (Applied Biosystems) with the following primers: *ETV1* promoter, 5-caaccacgtgaccaagaag-3 and 5-GCGCTCCGCTAGGAGATT-3; *ETV4* promoter, 5-cttctctcttttctctcggttc-3 and 5-CCAATCAGAATGTAGGGGTTG-3; *ETV5* promoter, 5-aagtgcttcactgactcagctaa-3 and 5-CATTGGCCAATCAGCACA-3. As a negative control we used a region of the *CDK1* promoter without known CBSs, amplified with primers 5-ggccttcaacgtatgaattagc-3 and 5-AGTTGGTATTGCACATAAGTCT-3.

In vitro DNA binding assays

EMSA experiments were performed using CIC protein fragments synthesized with the TNT T7 Quick Coupled Transcription/Translation system (Promega). For expression of His-tagged proteins, bacterial cultures were induced for 2 h with 1 mM IPTG and proteins purified using the Proteus IMAC Mini Sample kit. DNA probes were synthesized as complementary oligonucleotides leaving 5' GG overhangs, or amplified by PCR with primers carrying NotI restriction sites, subcloned, and released by NotI digestion. Probes were then end-labeled

using α-^{32}P-dCTP and Klenow Fragment, exo- (Thermo Scientific). The sequences of wild-type and mutant probes are shown in S1 Table.

Binding reactions were carried out in a total volume of 20 μl containing 60 mM Hepes pH 7.9, 20 mM Tris-HCl pH 7.9, 300 mM KCl, 5 mM EDTA, 5 mM DTT, 12% glycerol, 1 μg poly (dI-dC), 1 μg BSA, ~1 ng of DNA probe, and 1 μl of programmed or non-programmed (control) TNT lysate (or ~1 ng of bacterially expressed His-tagged protein). After incubation for 20 min on ice, protein-DNA complexes were separated on 5% non-denaturing polyacrylamide gels run in 0.5X TBE at 4˚C, and detected by autoradiography.

Supporting information

S1 Fig. Isolation of a CRISPR-Cas9-induced mutation in the C1 motif of CIC. Shown is a diagram of the targeted sequence indicating the protospacer and protospacer adjacent motif (PAM) elements. The predicted cleavage site of Cas9 is indicated by an arrowhead. A sequencing chromatogram of a PCR product amplified from a cic^{4} homozygous fly is shown below; note the loss of the sequence encoding the RQKL motif.
(TIF)

S2 Fig. The cic^{4} allele is a strong hypomorph. (A-C) Cuticles of embryos derived from females of the indicated genotypes. The cic^{1} allele is a strong hypomorphic mutation specifically affecting CIC function in the early embryo. cic^{Q474X} is a nonsense mutation upstream of the HMG-box coding region and behaves as a genetic null. *Df(3R)ED6027* is a deletion that removes the *cic* locus. Embryos from cic^{4}/cic^{1} females often exhibit small patches of cuticle with ventral denticles (arrowhead in A), indicating some residual differentiation of abdominal structures; in contrast, such denticles are never seen in embryos from cic^{Q474X}/cic^{1} or *Df(3R)ED6027*/cic^{1} females. (D-F) Representative wings from flies of the indicated genotypes. Note that cic^{4} homozygous mutant wings are less affected (e.g. show less ectopic vein material and blisters) than cic^{Q474X}/cic^{4} or *Df(3R)ED6027*/cic^{4} wings. Thus, cic^{4} is a weaker allele than cic^{Q474X} or *Df(3R)ED6027* in the two contexts examined.
(TIF)

S3 Fig. Subcellular localization of CIC constructs in human cells. (A-E") Confocal images of 293T cells transfected with the indicated GFP-tagged constructs and co-stained using anti-GFP antibody (A-E) and Hoechst 33342 (A'-E'). Control expression of GFP alone is shown in A'-A". Note that all CIC derivatives are localized to the nucleus.
(TIF)

S4 Fig. A minimal CIC protein composed of N2, HMG-box and C1 domains is functional in the early embryo. (A) Diagram of the HA-tagged Cic(N2-HMG-C1) derivative. The structural arrangement of the HMG-box and C1 domains is identical to that of construct 2 in Fig 5A. The N2 motif is described in ref. 37. (B) Expression of CIC(N2-HMG-C1)-HA in a blastoderm embryo stained with an anti-HA antibody. The protein was expressed using a transgene under the control of 5' and 3' *cic* genomic sequences [9,62]. (C, D) Maternal expression of CIC (N2-HMG-C1) significantly rescues the *cic* mutant (cic^{1}/cic^{Q474X}) phenotype. Note the presence of abdominal denticle belts in the rescued embryo (arrowheads). Panel D shows a control cic^{1}/cic^{Q474X} cuticle. (E, F) CIC(N2-HMG-C1) rescues the central band of *kni* mRNA expression in cic^{1}/cic^{Q474X} embryos. A control cic^{1}/cic^{Q474X} embryo lacking abdominal *kni* expression is shown in F.
(TIFF)

S1 Table. Sequences of probes used in EMSA experiments. The table lists the sequences of DNA probes used in Fig 5, with intact and mutated CIC sites highlighted in yellow. References describing the different CIC sites are also indicated.
(TIFF)

Acknowledgments

We thank A. Olza for *Drosophila* injections, L. Campos, B. Lim, Z. Paroush, F. Port, S. Shvartsman, A. Veraksa and J. Vilardell for discussions; and J. Jin, B. Edgar, C. MacKintosh, T. Nakamura and the Bloomington *Drosophila* Research Center for reagents, plasmids and strains.

Author Contributions

Conceptualization: MF LSC SGC MD MB GJ.

Funding acquisition: MB GJ.

Investigation: MF LSC LA NS.

Project administration: MB GJ.

Supervision: MD MB GJ.

Visualization: MF LSC MD GJ.

Writing – original draft: GJ.

Writing – review & editing: MF LSC SGC MD GJ.

References

1. Štros M, Launholt D, Grasser KD (2007) The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cell Mol Life Sci 64: 2590–2606. doi: 10.1007/s00018-007-7162-3 PMID: 17599239
2. Malarkey CS, Churchill MEA (2012) The high mobility group box: the ultimate utility player of a cell. Trends Biochem Sci 37: 553–562. doi: 10.1016/j.tibs.2012.09.003 PMID: 23153957
3. Kamachi Y, Kondoh H (2013) Sox proteins: regulators of cell fate specification and differentiation. Development 140: 4129–4144. doi: 10.1242/dev.091793 PMID: 24086078
4. Jiménez G, Shvartsman SY, Paroush Z (2012) The Capicua repressor—a general sensor of RTK signaling in development and disease. J Cell Sci 125: 1383–1391. doi: 10.1242/jcs.092965 PMID: 22526417
5. Jiménez G, Guichet A, Ephrussi A, Casanova J (2000) Relief of gene repression by Torso RTK signaling: role of *capicua* in *Drosophila* terminal and dorsoventral patterning. Genes Dev 14: 224–231. PMID: 10652276
6. Goff DJ, Nilson LA, Morisato D (2001) Establishment of dorsal-ventral polarity of the *Drosophila* egg requires *capicua* action in ovarian follicle cells. Development 128: 4553–4562. PMID: 11714680
7. Roch F, Jiménez G, Casanova J (2002) EGFR signalling inhibits Capicua-dependent repression during specification of *Drosophila* wing veins. Development 129: 993–1002. PMID: 11861482
8. Atkey MR, Boisclair Lachance JF, Walczak M, Rebello T, Nilson LA (2006) Capicua regulates follicle cell fate in the *Drosophila* ovary through repression of *mirror*. Development 133: 2115–2123. doi: 10.1242/dev.02369 PMID: 16672346
9. Astigarraga S, Grossman R, Díaz-Delfín J, Caelles C, Paroush Z, et al. (2007) A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. EMBO J 26: 668–677. doi: 10.1038/sj.emboj.7601532 PMID: 17255944
10. Tseng ASK, Tapon N, Kanda H, Cigizoglu S, Edelmann L, et al. (2007) Capicua regulates cell proliferation downstream of the Receptor Tyrosine Kinase/Ras signaling pathway. Curr Biol 8: 728–733. doi: 10.1016/j.cub.2007.03.023 PMID: 17398096

11. Löhr U, Chung HR, Beller M, Jäckle H (2009) Antagonistic action of Bicoid and the repressor Capicua determines the spatial limits of *Drosophila* head gene expression domains. Proc Natl Acad Sci USA 106: 21695–21700. doi: 10.1073/pnas.0910225106 PMID: 19959668

12. Ajuria L, Nieva C, Winkler C, Kuo D, Samper N, et al. (2011) Capicua DNA-binding sites are general response elements for RTK signaling in *Drosophila*. Development 138: 915–924. doi: 10.1242/dev.057729 PMID: 21270056

13. Lim B, Samper N, Lu H, Rushlow C, Jiménez G, et al. (2013) Kinetics of gene derepression by ERK signaling. Proc Natl Acad Sci USA 110: 10330–10335. doi: 10.1073/pnas.1303635110 PMID: 23733957

14. Jin Y, Ha N, Forés M, Xiang J, Gläßer C, et al. (2015) EGFR/Ras signaling controls *Drosophila* intestinal stem cell proliferation via Capicua-regulated genes. PLoS Genet 11: e1005634. doi: 10.1371/journal.pgen.1005634 PMID: 26683696

15. Dissanayake K, Toth R, Blakey J, Olsson O, Campbell DG, et al. (2011) ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription factors via the transcriptional repressor *capicúa*. Biochem J 433: 515–525. doi: 10.1042/BJ20101562 PMID: 21087211

16. Fryer JD, Yu P, Kang H, Mandel-Brehm C, Carter AN, et al. (2011) Exercise and genetic rescue of SCA1 via the transcriptional repressor Capicua. Science 334: 690–693. doi: 10.1126/science.1212673 PMID: 22053053

17. Lee Y, Fryer JD, Kang H, Crespo-Barreto J, Bowman AB, et al. (2011) ATXN1 protein family and CIC regulate extracellular matrix remodeling and lung alveolarization. Dev Cell 21: 746–757. doi: 10.1016/j.devcel.2011.08.017 PMID: 22014525

18. Kim E, Park S, Choi N, Lee J, Yoe J, et al. (2015) Deficiency of Capicua disrupts bile acid homeostasis. Sci Rep 5: 8272. doi: 10.1038/srep08272 PMID: 25653040

19. Lam YC, Bowman AB, Jafar-Nejad P, Lim J, Richman R, et al. (2006) ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. Cell 127: 1335–1347. doi: 10.1016/j.cell.2006.11.038 PMID: 17190598

20. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, et al. (2006) The consensus coding sequences of human breast and colorectal cancers. Science 314: 268–274. doi: 10.1126/science.1133427 PMID: 16959974

21. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, et al. (2012) Recurrent R-spondin fusions in colon cancer. Nature 488: 660–664. doi: 10.1038/nature11282 PMID: 22895193

22. Bettegowda C, Agrawal N, Jiao Y, Sausen M, Laura D, et al. (2011) Mutations in *CIC* and *FUBP1* contribute to human oligodendroglioma. Science 333: 1453–1455. doi: 10.1126/science.1210557 PMID: 21817013

23. Okimoto RA, Breitenbuecher F, Olivas VR, Wu W, Gini B, et al. (2017) Inactivation of Capicua drives cancer metastasis. Nat Genet 49: 87–96. doi: 10.1038/ng.3728 PMID: 27869830

24. Jiao Y, Killela PJ, Reitman ZJ, Rasheed BA, Heaphy CM, et al. (2012) Frequent *ATRX*, *CIC*, *FUBP1* and *IDH1* mutations refine the classification of malignant gliomas. Oncotarget 3: 709–722. doi: 10.18632/oncotarget.588 PMID: 22869205

25. Sahm F, Koelsche C, Meyer J, Pusch S, Lindenberg K, et al. (2012) *CIC* and *FUBP1* mutations in oligodendrogliomas, oligoastrocytomas and astrocytomas. Acta Neuropathol 123: 853–860. doi: 10.1007/s00401-012-0993-5 PMID: 22588899

26. Yip S, Butterfield YS, Morozova O, Chittaranjan S, Blough MD, et al. (2012) Concurrent *CIC* mutations, *IDH* mutations, and 1p/19q loss distinguish oligodendrogliomas from other cancers. J Pathol 226: 7–16. doi: 10.1002/path.2995 PMID: 22072542

27. Chan AKY, Pang JC-S, Chung NY-F, Li KKW, Poon WS, et al. (2014) Loss of CIC and FUBP1 expressions are potential markers of shorter time to recurrence in oligodendroglial tumors. Mod Pathol 27: 332–342. doi: 10.1038/modpathol.2013.165 PMID: 24030748

28. Chittaranjan S, Chan S, Yang C, Yang KC, Chen V, et al. (2014) Mutations in CIC and IDH1 cooperatively regulate 2-hydroxyglutarate levels and cell clonogenicity. Oncotarget 5: 7960–7979. doi: 10.18632/oncotarget.2401 PMID: 25277207

29. Padul V, Epari S, Moiyadi A, Shetty P, Shirsat NV (2015) ETV/Pea3 family transcription factor-encoding genes are overexpressed in *CIC*-mutant oligodendrogliomas. Genes Chromosomes Cancer 54: 725–733. doi: 10.1002/gcc.22283 PMID: 26357005

30. Gleize V, Alentorn A, Connen de Kérillis L, Labussière M, Nadaradjane AA, et al. (2015) *CIC* inactivating mutations identify aggressive subset of 1p19q codeleted gliomas. Ann Neurol 78: 355–374. doi: 10.1002/ana.24443 PMID: 26017892

31. Kawamura-Saito M, Yamazaki Y, Kaneko K, Kawaguchi N, Kanda H, et al. (2006) Fusion between *CIC* and *DUX4* up-regulates *PEA3* family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Hum Mol Genet 15: 2125–2137. doi: 10.1093/hmg/ddl136 PMID: 16717057

32. Yoshimoto M, Graham C, Chilton-MacNeill S, Lee E, Shago M, et al. (2009) Detailed cytogenetic and array analysis of pediatric primitive sarcomas reveals a recurrent *CIC-DUX4* fusion gene event. Cancer Genetics Cytogenet 195: 1–11. doi: 10.1016/j.cancergencyto.2009.06.015 PMID: 19837261

33. Graham C, Chilton-MacNeill S, Zielenska M, Somers GR (2012) The *CIC-DUX4* fusion transcript is present in a subgroup of pediatric primitive round cell sarcomas. Hum Pathol 43: 180–189. doi: 10.1016/j.humpath.2011.04.023 PMID: 21813156

34. Italiano A, Sung YS, Zhang L, Singer S, Maki RG, et al. (2012) High prevalence of *CIC* fusion with Double-Homeobox (DUX4) transcription factors in *EWSR1*-negative undifferentiated small blue round cell sarcomas. Genes Chromosomes Cancer 51: 207–218. doi: 10.1002/gcc.20945 PMID: 22072439

35. Choi EYK, Thomas DG, McHugh JB, Patel RM, Roulston D, et al. (2013) Undifferentiated small round cell sarcoma with t(4;19)(q35;q13.1) *CIC-DUX4* fusion: a novel highly aggressive soft tissue tumor with distinctive histopathology. Am J Surg Pathol 37: 1379–1386. doi: 10.1097/PAS.0b013e318297a57d PMID: 23887164

36. Machado I, Cruz J, Lavernia J, Rubio L, Campos J, et al. (2013) Superficial *EWSR1*-negative undifferentiated small round cell sarcoma with CIC/*DUX4* gene fusion: a new variant of Ewing-like tumors with locoregional lymph node metastasis. Virchows Arch 463: 837–842. doi: 10.1007/s00428-013-1499-9 PMID: 24213312

37. Forés M, Ajuria L, Samper N, Astigarraga S, Nieva C, et al. (2015) Origins of context-dependent gene repression by Capicua. PLoS Genet 11: e1004902. doi: 10.1371/journal.pgen.1004902 PMID: 25569482

38. Andreu MJ, Ajuria L, Samper N, González-Pérez E, Campuzano S, et al. (2012) EGFR-dependent downregulation of Capicua and the establishment of *Drosophila* dorsoventral polarity. Fly 6: 234–239. doi: 10.4161/fly.21160 PMID: 22878648

39. Andreu MJ, González-Pérez E, Ajuria L, Samper N, González-Crespo S, et al. (2012) Mirror represses *pipe* expression in follicle cells to initiate dorsoventral axis formation in *Drosophila*. Development 139: 1110–1114. doi: 10.1242/dev.076562 PMID: 22318229

40. Fuchs A, Cheung LS, Charbonnier E, Shvartsman SY, Pyrowolakis G (2012) Transcriptional interpretation of the EGF receptor signaling gradient. Proc Natl Acad Sci USA 109: 1572–1577. doi: 10.1073/pnas.1115190109 PMID: 22307613

41. Parkhurst SM, Bopp D, Ish-Horowicz D (1990) X: A ratio, the primary sex-determining in *Drosophila*, is transduced by helix-loop-helix proteins. Cell 63: 1179–1191. PMID: 2124516

42. Jiménez G, Paroush Z, Ish-Horowicz D (1997) Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. Genes Dev 11: 3072–3082. PMID: 9367988

43. Younger-Shepherd S, Vaessin H, Bier E, Jan LY, Jan YN (1992) *deadpan*, an essential pan-neural gene encoding an HLH protein, acts as a denominator in *Drosophila* sex determination. Cell 70: 911–922. PMID: 1525829

44. Goldstein RE, Jimenez G, Cook O, Gur D, Paroush Z (1999) Huckebein repressor activity in *Drosophila* terminal patterning is mediated by Groucho. Development 126: 3747–3755. PMID: 10433905

45. Goldstein RE, Cook O, Dinur T, Pisanté A, Karandikar UC, et al. (2005) An eh1-like motif in Odd-skipped mediates recruitment of Groucho and repression in vivo. Mol Cell Biol 25: 10711–10720. doi: 10.1128/MCB.25.24.10711-10720.2005 PMID: 16314497

46. Morán É, Jiménez G (2006) The Tailless nuclear receptor acts as a dedicated repressor in the early *Drosophila* embryo. Mol Cell Biol 26: 3446–3454. doi: 10.1128/MCB.26.9.3446-3454.2006 PMID: 16611987

47. Peirano RI, Wegner M (2000) The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences. Nucleic Acids Res 28: 3047–3055. PMID: 10931919

48. Bernard P, Tang P, Liu S, Dewing P, Harley VR, et al. (2003) Dimerization of SOX9 is required for chondrogenesis, but not for sex determination. Hum Mol Genet 12: 1755–1765. PMID: 12837698

49. Huang YH, Jankowski A, Cheah KSE, Prabhakar S, Jauch R (2015) SOXE transcription factors form selective dimers on non-compact DNA motifs through multifaceted interactions between dimerization and high-mobility group domains. Sci Rep 5: 10398–10398. doi: 10.1038/srep10398 PMID: 26013289

50. Atcha FA, Syed A, Wu B, Hoverter NP, Yokoyama NN, et al. (2007) A unique DNA binding domain converts T-cell factors into strong Wnt effectors. Mol Cell Biol 27: 8352–8363. doi: 10.1128/MCB.02132-06 PMID: 17893322

51. Chang MV, Chang JL, Gangopadhyay A, Shearer A, Cadigan KM (2008) Activation of Wingless targets requires bipartite recognition of DNA by TCF. Curr Biol 18: 1877–1881. doi: 10.1016/j.cub.2008.10.047 PMID: 19062282

52. Hoverter NP, Ting JH, Sundaresh S, Baldi P, Waterman ML (2012) A WNT/p21 circuit directed by the C-clamp, a sequence-specific DNA binding domain in TCFs. Mol Cell Biol 32: 3648–3662. doi: 10.1128/MCB.06769-11 PMID: 22778133

53. Hoverter NP, Zeller MD, McQuade MM, Garibaldi A, Busch A, et al. (2014) The TCF C-clamp DNA binding domain expands the Wnt transcriptome via alternative target recognition. Nucleic Acids Res 42: 13615–13632. doi: 10.1093/nar/gku1186 PMID: 25414359

54. Bhambhani C, Ravindranath AJ, Mentink RA, Chang MV, Betist MC, et al. (2014) Distinct DNA binding sites contribute to the TCF transcriptional switch in *C. elegans* and *Drosophila*. PLoS Genet 10: e1004133. doi: 10.1371/journal.pgen.1004133 PMID: 24516405

55. Ravindranath A, Cadigan KM (2014) Structure-function analysis of the C-clamp of TCF/Pangolin in Wnt/ß-catenin signaling. PLoS ONE 9: e86180. doi: 10.1371/journal.pone.0086180 PMID: 24465946

56. Archbold HC, Broussard C, Chang MV, Cadigan KM (2014) Bipartite recognition of DNA by TCF/Pangolin is remarkably flexible and contributes to transcriptional responsiveness and tissue specificity of Wingless signaling. PLoS Genet 10: e1004591. doi: 10.1371/journal.pgen.1004591 PMID: 25188465

57. ten Bosch JR, Benavides JA, Cline TW (2006) The TAGteam DNA motif controls the timing of *Drosophila* pre-blastoderm transcription. Development 133: 1967–1977. doi: 10.1242/dev.02373 PMID: 16624855

58. Port F, Chen HM, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. Proc Natl Acad Sci USA 111: E2967–E2976. doi: 10.1073/pnas.1405500111 PMID: 25002478

59. Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proc Natl Acad Sci USA 104: 3312–3317. doi: 10.1073/pnas.0611511104 PMID: 17360644

60. McNeill H, Yang CH, Brodsky M, Ungos J, Simon MA (1997) *mirror* encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the *Drosophila* eye. Genes Dev 11: 1073–1082. PMID: 9136934

61. Yeh E, Gustafson K, Boulianne GL (1995) Green fluorescent protein as a vital marker and reporter of gene expression in *Drosophila*. Proc Natl Acad Sci USA 92: 7036–7040. PMID: 7624365

62. Cinnamon E, Guri USA 92: 7036–7040. St Johnston D, Jiménez G, et al. (2004) Capicua integrates input from two maternal systems in Drosophila terminal patterning. EMBO J 23: 4571–4582. doi: 10.1038/sj.emboj.7600457 PMID: 15510215

63. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng 14: 529–532. PMID: 11579220

64. Chen X, Zaro JL, Shen WC (2013) Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev 65: 1357–1369. doi: 10.1016/j.addr.2012.09.039 PMID: 23026637

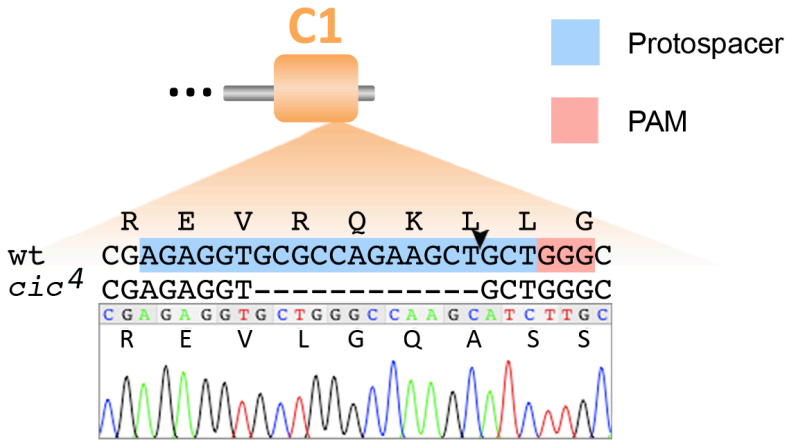
65. Maraver A, Fernandez-Marcos PJ, Herranz D, Cañamero M, Muñoz-Martin M, et al. (2012) Therapeutic effect of γ-secretase inhibition in *KrasG12V*-driven non-small cell lung carcinoma by derepression of DUSP1 and inhibition of ERK. Cancer Cell 22: 222–234. doi: 10.1016/j.ccr.2012.06.014 PMID: 22897852

66. Choi SH, Gearhart MD, Cui Z, Bosnakovski D, Kim M, et al. (2016) DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. Nucleic Acids Res 44: 5161–5173. doi: 10.1093/nar/gkw141 PMID: 26951377

67. de Celis JF, Bray S, Garcia-Bellido A (1997) Notch signalling regulates *veinlet* expression and establishes boundaries between veins and interveins in the *Drosophila* wing. Development 124: 1919–1928. PMID: 9169839

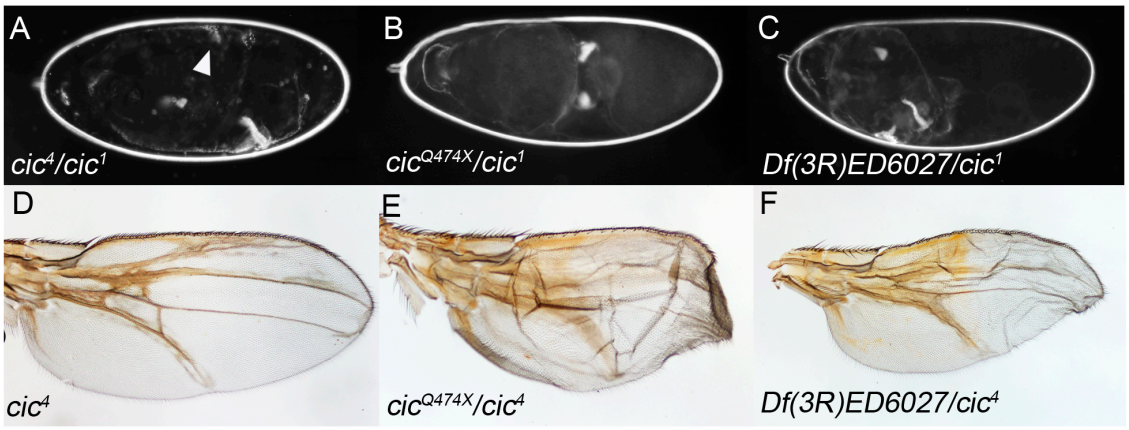
Supporting information

S1 Fig.



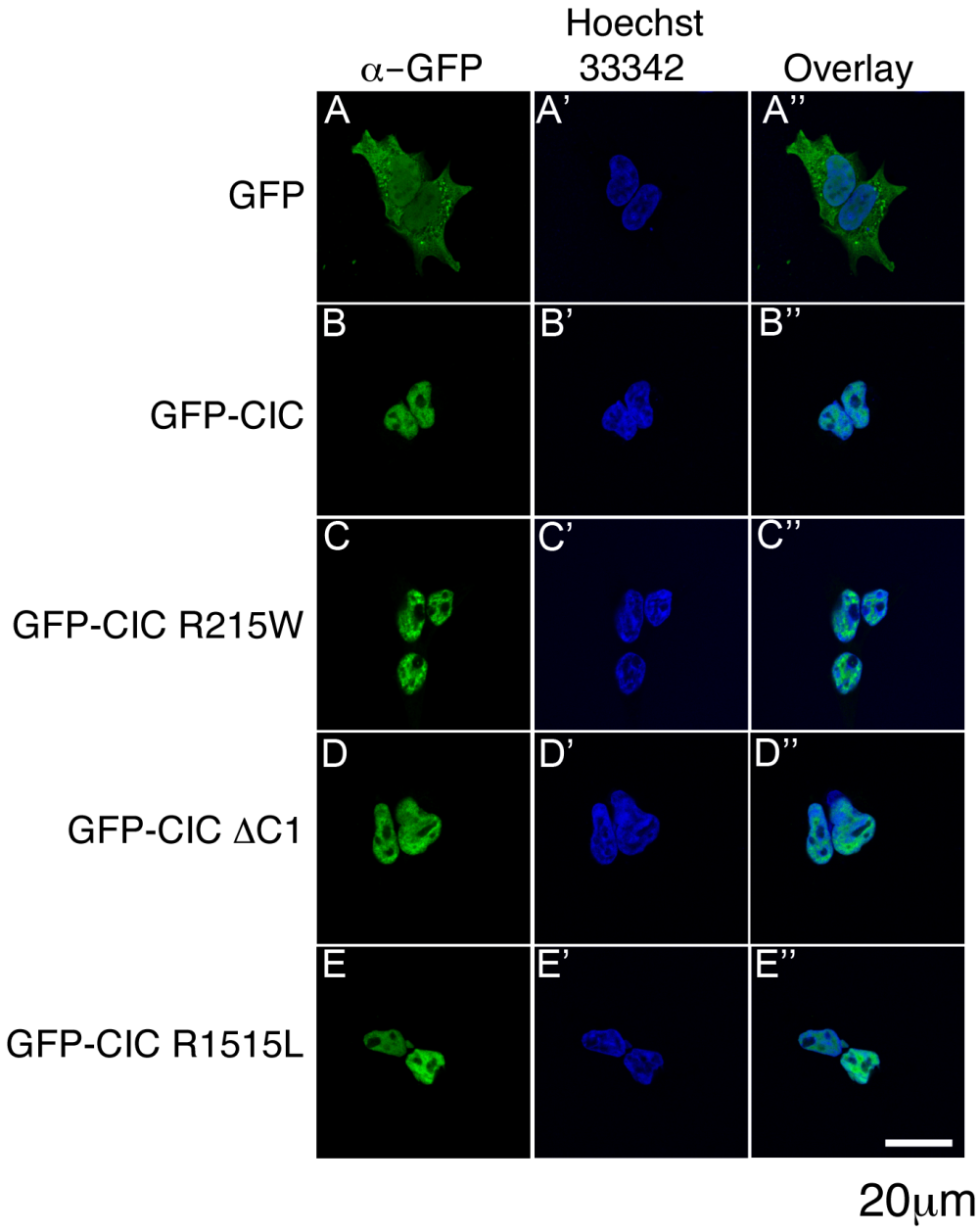
S1 Fig. Isolation of a CRISPR-Cas9-induced mutation in the C1 motif of CIC. Shown is a diagram of the targeted sequence indicating the protospacer and protospacer adjacent motif (PAM) elements. The predicted cleavage site of Cas9 is indicated by an arrowhead. A sequencing chromatogram of a PCR product amplified from a *cic*^{4} homozygous fly is shown below; note the loss of the sequence encoding the RQKL motif.

S2 Fig.



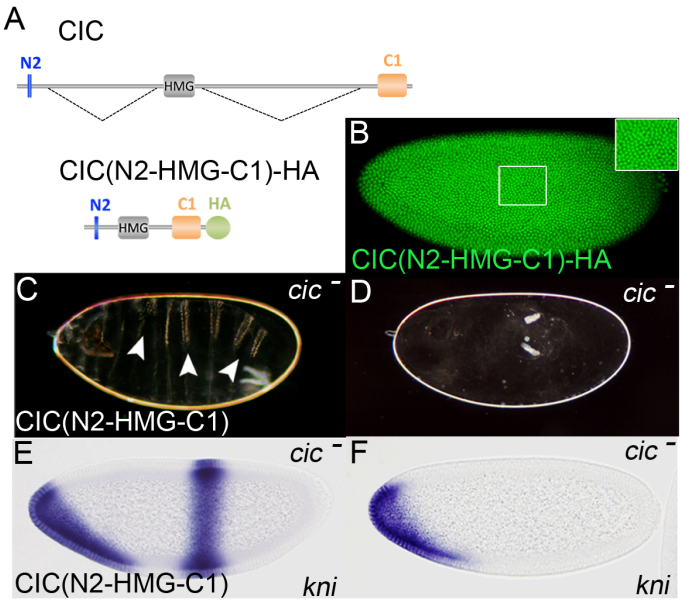
S2 Fig. The *cic*^{4} allele is a strong hypomorph. (A-C) Cuticles of embryos derived from females of the indicated genotypes. The *cic*^{1} allele is a strong hypomorphic mutation specifically affecting CIC function in the early embryo. *cic*^{Q474X} is a nonsense mutation upstream of the HMG-box coding region and behaves as a genetic null. *Df(3R)ED6027* is a deletion that removes the *cic* locus. Embryos from *cic*^{4}/*cic*^{1} females often exhibit small patches of cuticle with ventral denticles (arrowhead in A), indicating some residual differentiation of abdominal structures; in contrast, such denticles are never seen in embryos from *cic*^{Q474X}/*cic*^{1} or *Df(3R)ED6027/cic*^{1} females. (D-F) Representative wings from flies of the indicated genotypes. Note that *cic*^{4} homozygous mutant wings are less affected (e.g. show less ectopic vein material and blisters) than *cic*^{Q474X}/*cic*^{4} or *Df(3R)ED6027/cic*^{4} wings. Thus, *cic*^{4} is a weaker allele than *cic*^{Q474X} or *Df(3R)ED6027* in the two contexts examined.

S3 Fig.



S3 Fig. Subcellular localization of CIC constructs in human cells. (A-E") Confocal images of 293T cells transfected with the indicated GFP-tagged constructs and co-stained using anti-GFP antibody (A-E) and Hoechst 33342 (A'-E'). Control expression of GFP alone is shown in A'-A". Note that all CIC derivatives are localized to the nucleus.

S4 Fig.



S4 Fig. A minimal CIC protein composed of N2, HMG-box and C1 domains is functional in the early embryo. (A) Diagram of the HA-tagged Cic(N2-HMG-C1) derivative. The structural arrangement of the HMG-box and C1 domains is identical to that of construct 2 in Fig 5A. The N2 motif is described in ref. 37. (B) Expression of CIC(N2-HMG-C1)-HA in a blastoderm embryo stained with an anti-HA antibody. The protein was expressed using a transgene under the control of 5' and 3' *cic* genomic sequences [9,62]. (C, D) Maternal expression of CIC(N2-HMG-C1) significantly rescues the *cic* mutant (cic^{1}/cic^{Q474X}) phenotype. Note the presence of abdominal denticle belts in the rescued embryo (arrowheads). Panel D shows a control cic^{1}/cic^{Q474X} cuticle. (E, F) CIC(N2-HMG-C1) rescues the central band of *kni* mRNA expression in cic^{1}/cic^{Q474X} embryos. A control cic^{1}/cic^{Q474X} embryo lacking abdominal *kni* expression is shown in F.

S1 Table.

S1 Table. Sequences of probes used in EMSA experiments

Probe	Size	Sequence	Refs.
hkb 2xCBS	98bp	GGCCGCTCCCTGCTCACGGCAATCCCAGTTTATGAATGAATTTACT AAATGAGCTCAGTCGTGAATGAAATGTTCTCCGCGGACCGCGTTT TGCGGCC	[12]
hkb 2xCBSmut	98bp	GGCCGCTCCCTGCTCACGGCAATCCCAGTTTACACACGCATTTACT AAATGAGCTCAGTCGCACACGCAATGTTCTCCGCGGACCGCGTTT TGCGGCC	[12]
pnt 1xCBS	170bp	GGCCGCGATGGCCGTTTCCGTTCCAAGCACACAACACACAGACGC AGCCGAGAGAGCGCACGCACACACACGCACTCGCATCGTTTTATG AATGAATTCCATTTCGTTCCGCAAGAGGGTTTCCAGATTTACGAG TTCCGTTCCCACTTGTTGCGCTCTCTGGCGCGGCC	[14]
ETV4 2xCBS	159bp	GGCCGCGTCCTCGGCTTCTCTCTTTTTCTCTCGGTTCTTTATGAATG AACTCCAGCCGCGAGTTTTATTTTTTTTATGAATGGGAGCCGGAGC ACTGAATGAAGCCGGGAGCCAACCCCTACATTCTGATTGGCCCCC GCGGCTGGGAAACTCCGCGGCC	[15,17]
ETV5 1xCBS	132bp	GGCCGCGGCCCCTTTAAGTTCTTCACTGACTCAGCTAAGCTTCCCC CTCAGGTCGCGTTTTTTATGAATGAAAACGTCCTTACATTCATTA TATAAACAAGCCTGTGCTGATTGGCCAATGGCCCGCGGCC	[15,17,31]

S1 Table. Sequences of probes used in EMSA experiments. The table lists the sequences of DNA probes used in Fig 5, with intact and mutated CIC sites highlighted in yellow. References describing the different CIC sites are also indicated.

DISCUSIÓN

Desde su identificación hace más de 15 años, el factor represor Cic ha adquirido una importancia creciente como sensor general de las señales iniciadas por RTKs. En *Drosophila*, Cic actúa como represor de genes regulados por las vías de Torso y EGFR, las cuales ejercen su función, al menos en parte, mediante la inactivación de Cic (Atkey et al., 2006; Astigarraga et al., 2007; Andreu et al., 2012a). Además, diferentes estudios han demostrado la importancia de CIC durante el desarrollo del tejido pulmonar en ratón, así como su participación directa en enfermedades humanas como cáncer y ataxia espinocerebelar (Kawamura-Saito et al., 2006; Lam et al., 2006; Bettegowda et al., 2011; Dissanayake et al., 2011; Fryer et al., 2011; Lee et al., 2011). Todas estas funciones dependen de la capacidad de Cic de actuar como represor transcripcional mediante su unión directa a motivos T(C/G)AATG(A/G)A presentes en las regiones reguladoras de los genes diana (Kawamura-Saito et al., 2006; Ajuria et al., 2011; Lee et al., 2011). El mecanismo molecular que utiliza Cic para ejercer su actividad represora de la transcripción es aun ampliamente desconocido. Esta cuestión es importante para definir con precisión las actividades moleculares de Cic como regulador del desarrollo y otros procesos celulares. En relación con este aspecto, durante esta tesis se ha estudiado el mecanismo por el que Cic ejerce su actividad represora en distintos contextos: en humanos y en *Drosophila*.

En primer lugar, durante esta tesis se ha estudiado con detalle la relación funcional entre Cic y Gro en el embrión temprano de *Drosophila*, así como en otros tejidos de la mosca. También se han iniciado estudios para profundizar en el origen evolutivo de esta relación funcional.

En segundo lugar, hemos estudiado los mecanismos de regulación de Cic a través de las vías RTK y hemos explorado nuevas vías implicadas en la regulación de su actividad.

Por último, otra cuestión que quedaba hasta ahora sin resolver acerca del mecanismo de acción de Cic era la función del motivo C1, presente en todas las formas conocidas de Cic a lo largo de la evolución e imprescindible para la actividad de la proteína al menos en *Drosophila* (Astigarraga et al., 2007). En humanos se ha visto que este motivo es esencial tanto para la función supresora tumoral de la proteína como para la función oncogénica en sarcomas de tipo *Ewing-like*. Así que durante esta tesis también

se ha tratado de resolver la función molecular de este motivo para entender mejor el mecanismo de acción que utiliza Cic para ejercer la represión de sus genes diana.

1. Relación funcional entre Cic y Gro

Hace años que se conoce que tanto Cic como Gro participan en la represión de los genes *gap* terminales *tll* y *hkb* (Paroush et al., 1997; Jiménez et al., 2000; Cinnamon et al., 2008; Jennings et al., 2008). Además, se ha asociado la presencia de Gro al *enhancer* de *hkb* a través de ensayos de inmunoprecipitación de cromatina (Ajuria et al., 2011), pero se desconocía hasta ahora la relación molecular entre estos dos elementos y si esta relación se da en todos los contextos donde actúa Cic.

1.1. Mecanismo represor de Cic dependiente de contexto en el desarrollo de *Drosophila*

El hecho de que tanto Cic como Gro sean imprescindibles para la represión de *tll* y *hkb* no implica que ambas proteínas estén interaccionando de manera directa, sino que plantea dos posibles escenarios: el primero, dónde la interacción entre Cic y Gro sería directa o a través de una proteína adaptadora; y el segundo, dónde Gro estaría interaccionando con otro factor que reconocería los mismos *enhancers* que Cic. Para resolver esta cuestión, insertamos lugares de unión de Cic (CBSs) en un *enhancer* heterólogo. Los resultados mostraron que la represión de la expresión de este *enhancer* sintético en el embrión depende tanto de Cic como de Gro, sugiriendo que los CBSs son suficientes para mediar represión a través de Cic y de manera dependiente de Gro (Forés et al., 2015; Fig. 2) y demostrando que es Cic quien recluta a Gro a sus *enhancers*, al menos en el embrión temprano.

Por otro lado, analizamos la actividad represora de Cic en otros tejidos y estudiamos si el requerimiento de Gro se da por igual en todos los contextos. Sorprendentemente, encontramos que las actividades de Cic son dependientes de contexto, puesto que Cic puede actuar de manera dependiente o independiente de Gro en función del tejido. Por un lado, en el embrión temprano la función de Cic requiere del correpresor Gro para reprimir sus genes diana *tll* y *hkb.* Además, hay indicios de que otras funciones de Cic en el embrión también podrían requerir la actividad de Gro, ya que tanto pérdidas de función de *cic* como de *gro* causan la desrepresión de *ind* (Ajuria et al., 2011;

Helman et al., 2011) y de *zen* (Dubnicoff et al., 1997; Jiménez et al., 2000), aunque para esta última función se ha sugerido que es Dorsal quien recluta a Gro (Dubnicoff et al., 1997). En cambio, en otros tejidos Cic realiza su actividad represora sin el requerimiento de Gro. En el epitelio folicular del ovario y en el disco imaginal de ala, Cic reprime a *mirr* y *aos* respectivamente (Atkey et al., 2006; Ajuria et al., 2011; Andreu et al., 2012b), por lo que clones de perdida de función de *cic* resultan en la desrepresión de estos genes diana (Forés et al., 2015; Fig. 1). Sin embargo, clones de perdida de función de *gro* no afectan al patrón de expresión de *mirr* y *aos* (Forés et al., 2015; Fig. 1), lo cual es consistente con la observación de que pérdidas de función de *gro* no afectan a la expresión del gen *pipe* en el epitelio folicular (Technau et al., 2012). En conjunto, estos resultados indican que Cic requiere del correpresor Gro para reprimir en el embrión de *Drosophila,* pero no en otros momentos del desarrollo como en el epitelio folicular o el disco imaginal de ala.

1.2. El motivo N2 es un elemento represor dependiente de Gro

Hemos profundizado en las funciones de Cic dependientes de Gro y hemos visto que Cic contiene un motivo esencial para el reclutamiento de Gro al que hemos denominado N2, el cual es diferente a los motivos definidos hasta ahora de interacción con Gro (los motivos eh1 y WRPW).

El motivo N2 está altamente conservado en dípteros, es esencial para la capacidad represora de la proteína y tiene la particularidad de que su secuencia esta codificada al final del primer exón de la isoforma Cic-S. Curiosamente, la secuencia codificada al otro lado del intrón también está altamente conservada y es común para las isoformas Cic-S y Cic-L. Sorprendentemente, solo el motivo N2, presente únicamente en la isoforma corta, es vital para la función represora de la proteína en el embrión (Forés et al., 2015; Fig. 3).

Dado que el motivo N2 está altamente conservado en diferentes especies de dípteros y que resulta esencial para la función de la proteína en el embrión, estudiamos la posibilidad de que fuera un motivo esencial para el reclutamiento de Gro. Mostramos que el motivo N2 tiene una capacidad represora intrínseca, transferible y dependiente de Gro, ya que en un ensayo heterólogo, el motivo N2 requiere de la presencia de Gro

para reprimir, mientras otros motivos conservados de la proteína no son capaces de reprimir en el mismo ensayo (Forés et al., 2015; Fig. 4). Además el dominio N2 es substituible por otros motivos de interacción con Gro como el motivo eh1 en el contexto de la proteína Cic-S (Forés et al., 2015; Fig. 5).

Un aspecto interesante de la posible interacción del motivo N2 con Gro es el hecho de que se dé manera distinta a la de los otros motivos de unión a Gro conocidos. Gro adopta en solución una conformación llamada *β-propeller*, en la que los 7 dominios WD forman una estructura de 7 aspas alrededor de un poro central (Pickles et al., 2002). Estudios genéticos y moleculares han mostrado que tanto el motivo WRPW como el motivo eh1 interaccionan con Gro a través de este poro central, adoptando una conformación similar (Jennings et al., 2006). La mutación MB41 de Gro afecta al poro y hace que los motivos WRPW y eh1 sean incapaces de interaccionar con él, pero sin embargo mantiene la capacidad de interaccionar con Cic, puesto que la expresión de *kni* en estos mutantes es normal (Jennings et al., 2006). Aprovechamos las particularidades de la mutación MB41 para evaluar la capacidad represora de una forma de Cic donde el motivo N2 estaba substituido por el motivo eh1. La forma Cic^{eh1} es incapaz de reprimir en presencia de la mutación MB41, indicando que es el motivo N2 quien hace que Cic sea funcional en presencia de esta mutación, y que por lo tanto este motivo es el encargado de reclutar a Gro a sus genes diana.

El hecho que Cic resulte insensible a la mutación MB41 plantea dos posibilidades: que el motivo N2 reconozca a Gro a través de una región diferente al poro de *β-propeller*, o que lo haga a través del poro, pero adoptando una conformación distinta a la de los otros dos motivos, lo que lo hace insensible a la mutación. De hecho, ya se ha visto que hay ciertas mutaciones del poro, como la mutación MB31, que afectan de manera distinta a los motivos WRPW y eh1, ya que éstos adoptan conformaciones ligeramente distintas a través de él (Figura 13) (Jennings et al., 2006). Hay otra mutación de Gro, llamada E48, que también afecta al poro y está definida como un nulo de la proteína, ya que inactiva todas las funciones conocidas de Gro, incluida la regulación del sistema terminal (Delidakis et al., 1991; Jennings et al., 2006). No está claro si esta mutación desestabiliza de algún modo la proteína, pero en caso que no lo hiciera, se podría

sugerir que está afectando a la interacción con los motivos represores en el poro, incluido el N2 (Figura 13).

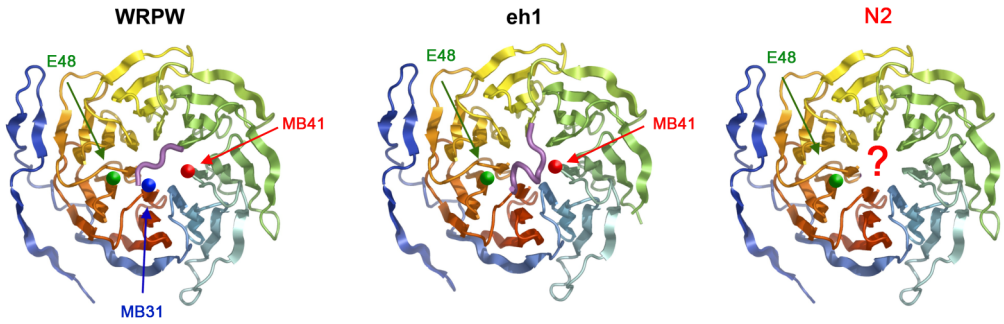


Figura 13. Interacción de Gro con los motivos represores. Modelos estructurales de la unión de Gro a diferentes motivos represores. Las mutaciones que afectan la unión en cada caso están indicadas. Si el motivo N2 interacciona con el poro de *β-propeller*, lo estará haciendo de un modo distinto al que lo hacen los motivos WRPW y eh1. (Adaptado de Jennings et al., 2006)

Por otro lado, se ha visto que el motivo WRPY de Runt se une al poro del mismo modo que WRPW a pesar de ser dos motivos ligeramente distintos (Jennings et al., 2006). Del mismo modo, el motivo N2 guarda cierta similitud con diferentes motivos eh1 (Figura 14), por lo que una posibilidad es que, si la interacción del N2 se produce en el poro, podría estar haciéndolo de una manera similar al motivo eh1. Un aminoácido clave para la interacción del eh1 con Gro es la fenilalanina inicial, la interacción de la cual se ve afectada por la mutación MB41. El anillo aromático de la cadena lateral de la fenilalanina del motivo eh1 adopta la misma posición que el anillo aromático del triptófano inicial del motivo WRPW. Así, parece que es esencial tener un anillo aromático en posiciones iniciales para la interacción. El motivo N2 carece de aminoácidos con cadenas laterales aromáticas por lo que, de interaccionar con el poro, su interacción se estaría dando de un modo distinto a como lo hace eh1 al no poder contactar los mismos residuos a través de él.

```
eh1   FxIxxIL
En    FSISNIL
Gsc   FSIDNIL
Odd   FTIDEIM
Bowl  FSIEDIM
N2    LYLQCLL
```

Figura 14. Conservación del motivo eh1 en diferentes proteínas. Secuencias de los motivos eh1: Engrailed (En), Gooscoid (Gsc), Oddskipped (Odd), Bowl (Bowl) y el motivo N2 de Cic. El motivo N2 guarda cierta similitud a los motivos eh1 aunque difiere en la fenilalanina en la posición inicial.

El hecho de que no hayamos sido capaces de detectar una unión directa entre Cic y Gro a través de diferentes ensayos (interacciones directas *in vitro* y ensayos de dos

híbridos), deja abierta la posibilidad de que la interacción entre Cic y Gro no sea directa. En este escenario, Cic estaría reclutando a una proteína adaptadora que sería la que se une con Gro directamente a través del poro o en otra región de su secuencia. Otra alternativa es que la unión entre Cic y Gro no sea lo suficientemente fuerte como para detectarse en ensayos de interacción *in vitro*. Estas dos posibilidades no son excluyentes, de hecho, las proteínas Runx se unen de manera débil a Gro a través de su motivo WRPY de manera que la interacción no se puede detectar en ensayos *in vitro* y además, esta interacción depende de proteínas accesorias *in vivo* (Aronson et al., 1997; Canon and Banerjee, 2003; Jennings et al., 2006).

1.3. Mecanismo de represión por Cic fuera del embrión

Los resultados obtenidos indican que el motivo N2 es el responsable de la represión mediada por Cic en el embrión temprano, pero, ¿cómo Cic reprime a sus genes diana en otros contextos? La isoforma Cic-S es suficiente para ejercer una respuesta transcripcional en las células foliculares y el disco imaginal de ala, así que si Cic está interaccionando con otro correpresor lo estará haciendo a través de algún motivo presente en esta isoforma (Astigarraga et al., 2007, y observaciones de nuestro laboratorio sin publicar). Una posibilidad es que el motivo N2 esté reclutando un correpresor distinto a Gro en el epitelio folicular y el disco imaginal de ala. Observaciones recientes indican que esto no es así, ya que la pérdida del motivo N2 en estos tejidos mantiene la capacidad represora de Cic (observaciones no publicadas). Por otro lado, experimentos en células en cultivo han mostrado que en mamíferos CIC requiere del correpresor ATXN1 o su factor relacionado ATXN1L para potenciar su actividad represora (Lam et al., 2006; Crespo-Barreto et al., 2010), y que la interacción entre estos dos factores se da entre el dominio AXH de ATXN1 y una región conservada de CIC, también presente en la isoforma Cic-S de *Drosophila* (Figura 2) (Lam et al., 2006; Kim et al., 2013). Se ha estudiado Atx-1, el ortólogo de ATXN1 en *Drosophila*, en modelos de patogénesis de ataxia espinocerebelar (Tsuda et al., 2005), y el análisis del interactoma de Cic en el embrión de *Drosophila* ha demostrado una interacción entre ambas proteínas (Yang et al., 2016). Por lo tanto, un posible cofactor en los tejidos donde Cic reprime sin la necesidad de Gro podría ser Atx-1. Resultados preliminares, realizados como complemento a este trabajo, indican que es improbable que Atx-1 sea

un correpresor de Cic en *Drosophila*, puesto que alelos de pérdida de función de *Atx-1* generados por la técnica CRISPR-Cas9, mantienen normal la polaridad DV del embrión y no muestran fenotipo de venación ectópica en el ala, por lo que los genes diana de Cic no parecen estar afectados en estos tejidos (resultados sin publicar).

Cic no es la única proteína con mecanismos de represión dependientes e independientes de Gro. Se ha visto que Runt también tiene un comportamiento dual, ya que puede reprimir algunos de sus genes diana como *even-skipped* (*eve*) o *hairy* a través de su interacción con Gro, pero sin embargo reprime *engrailed* (*en*) sin la necesidad del correpresor. En Runt no se han encontrado motivos represores adicionales por lo que se ha propuesto la posibilidad que cuando reprime independientemente de Gro lo haga a través de un mecanismo pasivo, compitiendo con activadores por un mismo sitio de unión (Aronson et al., 1997). Hemos analizado la secuencia de Cic, y tampoco hemos encontrado motivos represores adicionales al N2 ni al motivo de interacción con Atx-1, por lo que no podemos descartar que esté ocurriendo un mecanismo de represión pasiva similar en los tejidos donde Cic actúa independientemente de Gro.

1.4. Origen del motivo N2 y la isoforma Cic-S en dípteros

Otro aspecto que hemos estudiado es el origen evolutivo del motivo N2. Hemos visto que el motivo N2 se encuentra conservado en todas las especies de dípteros analizadas, pero no en otros insectos como lepidópteros, coleópteros, himenópteros, hemípteros ni tampoco en el resto de metazoos. Por lo tanto, la explicación más sencilla de esta distribución filogenética es que el motivo N2 se originó en los dípteros primitivos, hace aproximadamente 250 millones de años (Wiegmann et al., 2011) (Forés et al., 2015; Fig. 5). Además, dado que el motivo N2 se encuentra codificado en el primer exón de un nuevo tránscrito *cic-S* específico de dípteros, este motivo probablemente surgió mediante inserción de nuevas secuencias o expansión de una región intrónica del tránscrito ancestral *cic-L*, dando lugar a un nuevo exón y a una nueva región promotora con la subsiguiente evolución del motivo funcional N2 por sucesivas mutaciones puntuales (Forés et al., 2015; Fig. 6). De hecho, se ha sugerido que los motivos peptídicos cortos de interacción con Gro como WRPW y eh1 pueden

haber aparecido fácilmente mediante mutaciones puntuales durante la evolución (Hittinger and Carroll, 2008).

Los datos disponibles no nos permiten definir con precisión los cambios e innovaciones que desde el punto de vista del desarrollo supuso la adquisición del motivo N2 para la actividad represora de Cic-S, ya que se desconocen cuestiones básicas como la función molecular de la forma ancestral Cic-L, de manera que no sabemos cómo esta isoforma podría actuar como represor. Únicamente sabemos que esta forma ancestral no es capaz de substituir la función de la forma Cic-S para reprimir los genes terminales del embrión a no ser que incluya el motivo N2, lo cual confirma la actividad esencial del motivo N2 en el embrión de *Drosophila* (Forés et al., 2015; Fig. 6).

Una hipótesis de las ventajas moleculares que el motivo N2 le pudo conferir a los dípteros se puede encontrar en las dinámicas de represión diferenciales que utilizan los insectos. Es posible que la interacción con Gro a través del N2 haya generado un mecanismo de represión más robusto que permita a Cic reprimir de una manera más rápida en los embriones de dípteros en comparación con otras especies de insectos. En insectos no dípteros *tll* se induce en los polos del embrión por *orthodenticle* (*otd*) en *Nasonia* y por Torso en *Tribolium* (Schröder et al., 2000; Schoppmeier and Schröder, 2005; Lynch et al., 2006; Wilson and Dearden, 2009). Un estudio reciente, ha vinculado la correcta expresión de *tll* en el polo con la vía de Torso y con la función de Cic en el embrión de *Tribolium*. En este artículo, los autores han mostrado que, en pérdidas de función de la vía la expresión de *tll* desaparece, mientras que en pérdidas de función de *cic* éste se desreprime, indicando que la desaparición de *tll* en el polo en las pérdidas de función de la vía podrían deberse a ganancias de función de Cic (Pridöhl et al., 2017). Cómo Cic puede reprimir en el embrión de esta especie sin hacerlo a través de Gro es un misterio. *Tribolium,* al igual que el resto de insectos no dípteros, solo expresa una forma de Cic, la cual equivaldría a la forma ancestral Cic-L de dípteros (a no ser que haya generado otras formas cortas independientes de la forma corta de *Drosophila*), de manera que no contiene el motivo N2. Una posible explicación de la represión diferencial de Cic en las distintas especies de insectos se pueda encontrar en las diferentes velocidades que se requieren para el proceso de la embriogénesis. En *Drosophila*, el establecimiento del sistema terminal es un proceso que se establece

cuando el embrión está en fase de blastodermo: entre 1 y 3 horas desde que se deposita el huevo, el patrón de expresión de *tll* y *hkb* ya ha sido determinado. El coleóptero *Tribolium castaneum*, es un insecto de banda germinal corta, en el que la expresión de *tll* determina la zona de crecimiento, donde los segmentos son añadidos secuencialmente en el polo posterior del embrión (Revisado en Schröder et al., 2008). En este insecto, la especificación del patrón de expresión *tll* también se da en la etapa de blastodermo, que en *Tribolium* comprende el embrión de entre 12 y 24 horas de edad (El-Sherif et al., 2012). Así, mientras que en *Tribolium* el proceso de determinación de la expresión de *tll* dura unas 12 horas, en *Drosophila* este proceso se da en menos de 2 horas, requiriendo un mecanismo de represión rápido y robusto, que parece que se ha logrado mediante la interacción con Gro (Figura 15). De este modo, parece que la aparición de la función de Torso en los polos del embrión de *Tribolium* permitió a Cic empezar a regular el sistema terminal, diferenciándose del mecanismo empleado en himenópteros. La aparición de la forma Cic-S y el N2 permitió la interacción con Gro lo que conllevó una represión más rápida y efectiva en los embriones de los dípteros.

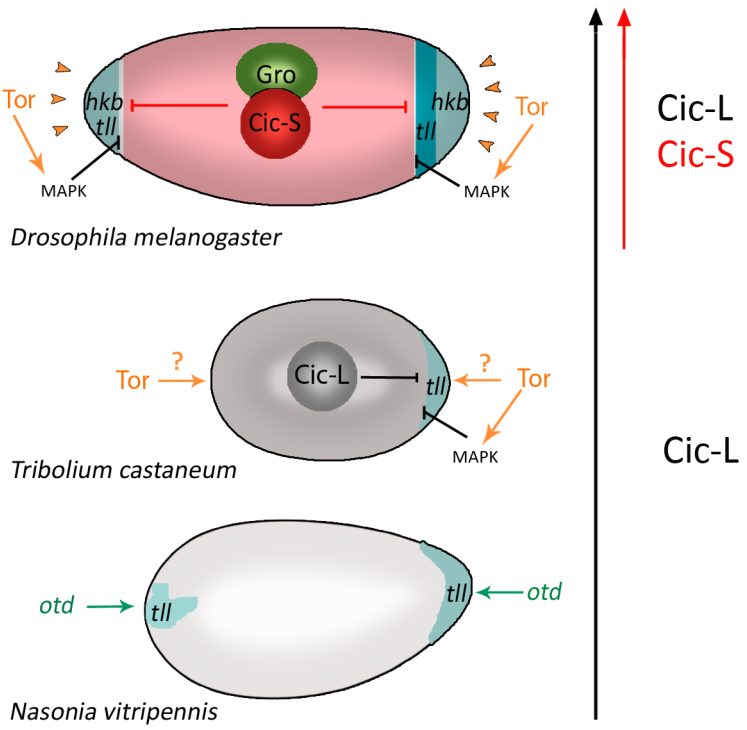


Figura 15. Modelo de la evolución del papel de Cic en la represión de los genes terminales. En *Nasonia*, la expresión de *tll* está controlada por *otd* y no hay datos que impliquen a Cic en este proceso. En el coleóptero *Tribolium*, se detecta por primera vez un control de la expresión de *tll* por parte de Cic y la vía de Torso. Esta función estaría ejecutada por la forma Cic-L. Se desconoce si la vía tiene algún mecanismo de inducción directa de la expresión de *tll*. En *Drosophila*, la aparición de la isoforma Cic-S y el motivo N2 ha transformado la represión de los genes terminales en un proceso dependiente de Gro.

Pero, ¿de qué manera la nueva interacción con Gro ha permitido a Cic ser un represor más robusto? Se ha visto que en el embrión de *Drosophila*, Cic actúa como un

morfógeno, de modo que presenta un gradiente de expresión donde el pico más alto corresponde a las regiones centrales del embrión y se reduce de forma gradual hasta eliminarse en los polos en respuesta a la señal de Torso (Astigarraga et al., 2007). De una manera similar, Gro también está fosforilado y funcionalmente inactivado en respuesta a la señal de Torso (Cinnamon et al., 2008; Helman et al., 2011). La combinación de gradientes de expresión de estas dos proteínas ha podido establecer el umbral necesario para reprimir cada uno de los genes diana. De este modo, los bordes de expresión de *tll* y *hkb* están regulados por los gradientes de Cic y Gro, y ligeros cambios en la concentración de Gro pueden modificar la concentración necesaria de Cic para reprimir (Astigarraga et al., 2007; Ajuria et al., 2011; Turki-Judeh and Courey, 2012). Estos gradientes superpuestos aseguran una represión más eficiente, ya que ante posibles fallos en alguno de los dos gradientes el mecanismo asegura una correcta represión al estar dentro de un umbral de "seguridad".

Existen pocos ejemplos en biología de casos similares, donde se revela el origen de una diferencia estructural y funcional dentro de una familia de reguladores del desarrollo. Un caso parecido es la aparición del motivo poly-Ala en la familia de reguladores Hox, la aparición del cual surgió durante la aparición de los hexápodos basales. En este caso, la evolución del dominio pudo facilitar la diversificación morfológica de los segmentos torácicos y abdominales posteriores característicos de los insectos modernos (Galant and Carroll, 2002).

Un aspecto que queda sin resolver es el mecanismo de acción que utiliza Gro para mediar represión. Uno de los modelos propuestos para el mecanismo de acción de Gro es que éste es reclutado a ADN a través de factores de transcripción, donde interacciona con la histona deacetilasa Rpd3, la cual modifica el estado de la cromatina para mantener una estructura transcripcionalmente silenciada (Chen et al., 1999; Chen and Courey, 2000). Se ha sugerido que Rpd3 se requiere solo para algunas funciones de Gro y, de hecho, es poco probable que la interacción de Cic con Gro reclute Rpd3, puesto que en pérdidas de función para *rpd3* en el embrión de *Drosophila* se observa un fenotipo de defectos de la segmentación tipo *pair-rule* que no corresponde con un fenotipo de perdida de función de *cic* (Chen et al., 1999).

Por último, resulta interesante señalar que en vertebrados existen isoformas CIC-S y CIC-L, y que la forma CIC-L guarda amplia similitud estructural con la isoforma correspondiente en insectos. La forma corta, en cambio, es diferente a la forma Cic-S de *Drosophila*, faltándole el motivo N2. Según nuestros análisis, esta nueva forma corta se originó con la aparición de los primeros vertebrados y también se podría haber creado por inserción de nuevas secuencias y expansión de la región interna del tránscrito de la forma larga, generando una isoforma diferente a las existentes (observaciones sin publicar). Parece, de este modo, improbable que en humanos CIC ejerza su función represora a través de TLE, el homólogo de Gro en humanos, ya que carece del motivo N2, a no ser que haya adquirido motivos de interacción con TLE distintos al N2. Nuestros estudios preliminares indican que, tal como ocurre en dípteros, el primer intrón de la isoforma CIC-S está muy conservado, por lo que sería interesante estudiar en futuros trabajos, la función diferencial y las innovaciones que esta nueva forma CIC-S puede haber aportado a los vertebrados.

2. Mecanismos de regulación de la actividad de Cic

Como ya se ha comentado anteriormente, la actividad de Cic está controlada por diferentes vías RTK tanto en *Drosophila* como en humanos.

En humanos, la regulación de CIC por la vía tiene un papel importante en el control de diferentes enfermedades. Por un lado, estudios en ratón han demostrado que en algunos casos de ataxia espinocerebelar, CIC se une de una manera más eficiente a sus genes diana causando una hiperrepresión. Este efecto se ve reducido disminuyendo los niveles de CIC o sometiendo a los ratones a ejercicio físico, lo cual aumenta la señalización EGFR e inhibe la actividad de CIC (Fryer et al., 2011). Por otro lado, se ha visto que mutaciones en *CIC* causan resistencias a tratamientos con inhibidores de la señal Ras/MAPK en tipos de cáncer donde la señalización por la vía está aumentada (Liao et al., 2017; Wang et al., 2017). Estas evidencias resaltan la importancia de conocer el mecanismo por el que las vías RTK controlan la expresión de la proteína CIC.

En *Drosophila*, la inactivación de Cic por las vías RTK es igualmente importante para regular los procesos en los que Cic está implicado. En esta especie, la interacción entre

MAPK y Cic se da en el motivo C2 para regular la actividad de la proteína, pero se desconoce si hay mecanismos de regulación adicionales (Astigarraga et al., 2007; Andreu et al., 2012a).

Para estudiar el efecto que tiene la vía sobre el motivo C2, se puede desacoplar a Cic de la señalización RTK mutando este motivo, lo que crea un represor dominante insensible a la vía. Los efectos de esta mutación solo se han estudiado utilizando transgenes de expresión o sobre-expresión de la forma Cic-S (Astigarraga et al., 2007; Andreu et al., 2012a), pero no en el contexto del locus endógeno de *cic* (el cual codifica para las isoformas Cic-L y Cic-S). Los transgenes $cic^{\Delta C2}$ causan fenotipos de ganancia de función en el embrión y en las células foliculares del ovario, lo cual afecta el establecimiento DV del embrión y la formación de apéndices respiratorios. Los adultos son viables, aunque la construcción causa esterilidad en las hembras al interferir con la función materna de Cic-S en el embrión temprano (Astigarraga et al., 2007).

2.1. El motivo C2 es importante para las isoformas Cic-S y Cic-L en *Drosophila*

Para analizar la regulación de Cic por la vía, estudiamos el efecto de mutar el motivo C2 en el locus endógeno de *cic* por la técnica CRISPR-Cas9. Mientras que se obtuvieron mutaciones relativamente grandes (de hasta de 20 pb) que alteraban el marco de lectura y generaban alelos de pérdida de función, resultó imposible aislar un alelo donde se delecionara el motivo C2 completo para crear un alelo de ganancia de función. Sin embargo, conseguimos aislar un alelo viable que afecta al motivo C2 (llamado cic^{3}), al que le falta un solo residuo (Forés et al., 2017a; Fig. 2). Estos resultados parecen indicar que este alelo debe mantener cierta capacidad de regulación por la vía y que deleciones más grandes del motivo podrían abolir completamente la unión y causar letalidad, ya que mutaciones mayores y en pauta no han podido ser recuperadas en el ensayo.

Deleciones completas del motivo C2 son viables en un transgén de la forma Cic-S, pero no en una mutación endógena. Esto plantea 2 posibilidades: que la letalidad venga dada por eliminar el motivo C2 afectando las 2 isoformas y que esto tenga un efecto aditivo; o que una deleción completa del dominio C2 estuviese afectando de manera

sustancial a Cic-L, la cual puede estar inactivada por la señal RTK a través del motivo C2. ¿Cuál es la función de Cic-L por debajo de la vía? Sabemos que Cic-L tiene un papel esencial en oogénesis y que hembras portadoras de alelos hipomorfos de *cic-L* dan lugar a embriones con fenotipos bicaudales (Rittenhouse and Berg, 1995). En este contexto desconocemos si Cic está regulado por las vías RTK. El hecho que no hayamos recuperado mutantes con el motivo C2 eliminado, incluso cuando la mutagénesis por CRISPR se ha generado en machos, sugiere que si la letalidad es debida a que el mutante está afectando significativamente la función de Cic-L, las funciones afectadas serían independientes al proceso de oogénesis.

Por otro lado, el mutante aislado cic^3, de acuerdo con lo esperado, resultó en una ganancia de función visible en diferentes sistemas (Forés et al., 2017a; Fig. 3), los cuales no se comentarán en detalle en la discusión puesto que ya están descritos en el artículo. De todos modos, resulta interesante añadir que también se observó una alteración en el patrón DV del embrión, ya que los embriones depositados por hembras cic^3 mostraron la banda de expresión de *twi* ensanchada, posiblemente debido a la reducción del dominio de expresión de *mirr* en el ovario y la consiguiente expansión de la expresión de *pipe* (observaciones no publicadas).

Otro aspecto interesante a destacar es el modo en que Cic^3 regula la formación de los apéndices. Como hemos mostrado en la publicación, los huevos depositados por hembras cic^3 mostraron una fusión de los apéndices respiratorios. No se sabe del todo bien que genes está regulando Cic en el proceso de la regulación del espacio entre apéndices, pero un posible candidato es *pnt*. En un estudio llevado a cabo en paralelo durante el desarrollo de esta tesis se identificó la relación entre Cic y Pnt como efectores por debajo de la vía EGFR en células madre intestinales (ISC) de *Drosophila* (Jin et al., 2015, ver Anexo). En este estudio comprobamos que Cic se une a las regiones reguladoras de *pnt* para reprimir su expresión. El hecho de que Cic esté controlando la expresión de *pnt* en ISCs, abre la posibilidad que esta relación esté conservada en otros procesos, como en el establecimiento de la línea media dorsal en el epitelio folicular. Mientras que en ISCs Cic estaría reprimiendo la expresión de *pnt* en el proceso de regeneración del epitelio intestinal, en el epitelio folicular del ovario,

esta relación funcional estaría regulando la organización de subpoblaciones de células para establecer una correcta posición de los apéndices dorsales.

2.2. La fosforilación por Minibrain y MAPK tiene efectos aditivos en la regulación de la actividad de Cic

Además de la regulación de Cic por MAPK, un trabajo en colaboración reciente ha mostrado que Cic puede estar regulado adicionalmente por la quinasa Minibrain (Mnb) y la proteína adaptadora Wings apart (Wap) (Yang et al., 2016). Mnb y Wap se unen a la región N-terminal de Cic (a una zona por lo tanto distinta del motivo C2) y fosforilan la proteína (Yang et al., 2016; Fig. 1,2), limitando la capacidad represora de Cic en tejidos como el disco imaginal de ala (Yang et al., 2016; Fig. 3).

Un aspecto que decidimos estudiar fue la relación entre la regulación de Cic por MAPK y Mnb. En el disco imaginal de ala, la función de Cic está regulada por MAPK (Roch et al., 2002), lo que es consistente con que un 30% de los adultos cic^3 presenten una pérdida parcial del tejido venoso. Sin embargo, la fosforilación por Mnb parece operar también en este tejido (Yang et al., 2016).

Una posibilidad es que el control por MAPK sea dependiente de Mnb. Se ha visto que DYRK1A, el homólogo de Mnb en humanos, aumenta la señalización de la vía MAPK formando un complejo con Ras, B-Raf y MEK (Kelly and Rahmani, 2005). Podría ser que en el disco imaginal de ala estuviera sucediendo algo parecido, y que Mnb tuviese un papel dual: fosforilar a Cic de manera directa, y contribuir a la fosforilación de Cic aumentando la señal MAPK de manera indirecta. Sin embargo, la depleción de Mnb no afecta a los niveles de MAPK fosforilada, por lo que descartamos la opción de que Mnb aumente la señal MAPK.

Consecuentemente, favorecemos un modelo en el que la regulación de Cic por MAPK y Mnb sean procesos independientes. En este escenario, Mnb podría estar ejerciendo un control basal o disminuyendo los niveles de Cic de manera constitutiva en las situaciones donde no hay señal MAPK. Para analizar esta posibilidad, se debería hacer un análisis más detallado con alelos mutantes de *mnb* o mediante análisis clonales. En favor de este modelo, hemos visto que MAPK y Mnb tienen un papel aditivo en la regulación de Cic, ya que cuando se elimina la función de Mnb en el ala, parte del

tejido venoso no se especifica, y este fenotipo se ve aumentado cuando se combina con el alelo *cic*^{3} (Yang et al., 2016; Fig. 6). Esta actividad aditiva controla la especificación del patrón de venación del ala, pero no parece operar en todas las células de igual modo. En las células que especifican la vena L5 en el disco imaginal de ala, ambas vías se requieren para inhibir Cic, en cambio en las células que especificaran las venas L2 y L3 la principal señal que inhibe a Cic es la de MAPK (Figura 16).

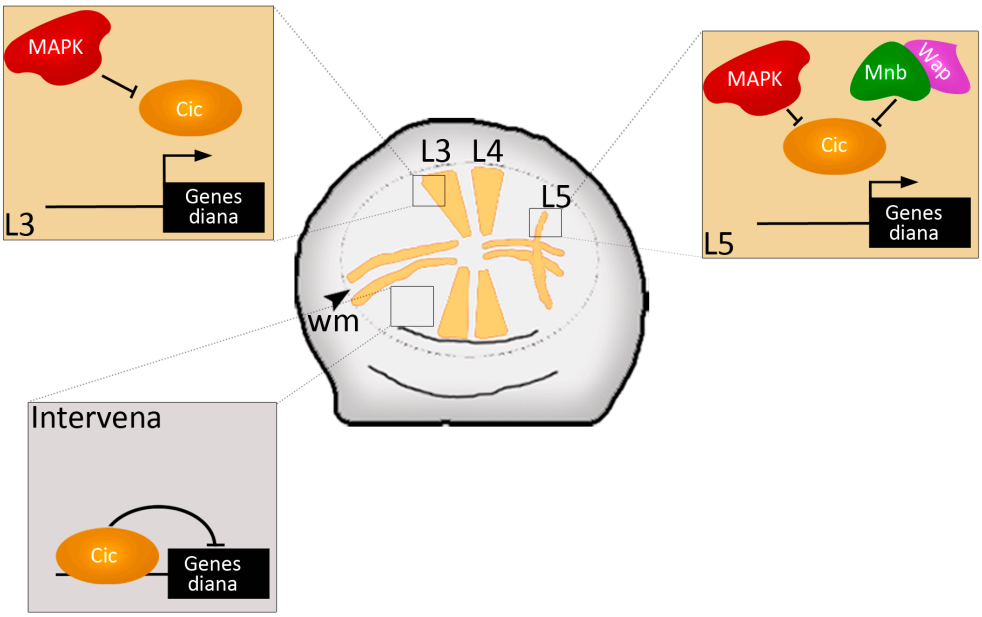


Figura 16. Representación de la regulación de Cic por MAPK y Mnb/Wap en el disco imaginal de ala. En las células que dan lugar a la vena L3, Cic se encuentra regulado mayoritariamente por MAPK, mientras que en las células que dan lugar a la vena L5, Cic está regulado de manera aditiva por MAPK y Mnb/Wap. Tanto en las venas L3 como L5, Cic está inactivado permitiendo la expresión de sus genes diana. En las regiones intervena, en cambio, Cic no está inactivado por ninguna de las vías de señalización por lo que actúa como represor de sus genes diana. Wm: *wing margin* (margen de la vena).

Aún queda por resolver el lugar exacto donde se da la interacción entre Mnb y Cic, y si este lugar está conservado a lo largo de la evolución en diferentes especies. La interacción entre Mnb y Wap está conservada en humanos (Tejedor et al., 1995; Skurat and Dietrich, 2004; Degoutin et al., 2013), así que sería interesante estudiar si la interacción entre los tres miembros del complejo también lo está. Además, en oligodendroglioma se han encontrado mutaciones en *CIC* (Bettegowda et al., 2011; Jiao et al., 2012; Sahm et al., 2012; Yip et al., 2012), así como una expresión elevada de DYRK1A (Pozo et al., 2013), abriendo la posibilidad que exista una relación funcional entre ambos factores en esta enfermedad.

Finalmente, otra cuestión importante respecto al control de Cic por MAPK es si éste se está dando a través de un único motivo o mediante mecanismos moleculares complementarios. En ensayos *in vitro* con la proteína CIC humana, se ha visto que MAPK se une a una región distinta al motivo C2 (Figura 2) (Futran et al., 2015). Esta región esta moderadamente conservada en *Drosophila*, por lo que, en esta especie, el motivo podría tener alguna capacidad residual de reclutar MAPK e inactivar la proteína. En caso de que esto fuera así, sería especialmente interesante analizar el efecto de delecionar esta región en el ala, porque a pesar de que hemos visto que MAPK y Mnb tienen efectos aditivos en este tejido, la disminución de la inactivación de Cic por estas dos vías no elimina completamente la formación de venas, lo que podría significar que la función de Cic no se ve comprometida del todo.

3. Mecanismo de unión de Cic a ADN

Cic contiene un dominio conservado HMG-box el cual se creía responsable de la unión de Cic al ADN, aunque las uniones detectadas *in vitro* utilizando este motivo son relativamente débiles (Jiménez et al., 2000; Kawamura-Saito et al., 2006; Ajuria et al., 2011). Además, se ha establecido la secuencia exacta a la que Cic se une en las regiones reguladoras de sus genes diana: se trata de un octámero con secuencia T(G/C)AATG(A/G)A tanto en *Drosophila* como en mamíferos (Kawamura-Saito et al., 2006; Ajuria et al., 2011; Lee et al., 2011).

Notablemente, se ha visto que el motivo HMG-box de CIC está especialmente afectado por mutaciones inactivantes en casos de oligodendroglioma y otros tipos de cáncer, lo cual concuerda con la función esencial de este dominio en la proteína para reconocer los genes diana. Curiosamente, hay otro motivo que se ha visto frecuentemente mutado en este mismo tipo de cáncer cerebral, que es el motivo C1 (Bettegowda et al., 2011; Jiao et al., 2012; Yip et al., 2012; Gleize et al., 2015; Padul et al., 2015). Ante estos resultados parece lógico pensar que tanto la HMG-box como el motivo C1 son imprescindibles para la función de CIC como supresor tumoral. ¿Pero el motivo C1 es también importante para la función oncogénica de CIC? En los casos en que CIC actúa como un oncogén (los casos de fusiones *CIC-DUX4* causantes de sarcomas de tipo *Ewing-like*), el motivo C1 parece también tener una función esencial, ya que las

translocaciones cromosómicas asociadas a esta fusión se dan en diferentes puntos de la secuencia de *CIC*, pero mayormente conservan el motivo C1 (Forés et al., 2017b; Fig. 1). En *Drosophila*, este motivo tiene funciones imprescindibles para mediar represión, ya que transgenes que carecen de él son incapaces de rescatar la función de la proteína en el embrión en fondos mutantes para *cic* (Astigarraga et al., 2007).

En este último trabajo, hemos tratado de esclarecer la función molecular, hasta ahora desconocida del motivo C1, el cual está presente en todas las formas conocidas de Cic, así como lo está la HMG-box.

3.1. El motivo C1 es esencial para las funciones de Cic como represor de sus genes diana en *Drosophila* y en humanos

Para conocer el requerimiento del motivo C1 para las funciones de Cic en *Drosophila*, utilizamos la técnica CRISPR-Cas9 con el objetivo de aislar un alelo mutante en el motivo C1. Este alelo, al que hemos llamado cic^{4}, causa una disminución de la función de Cic en diferentes contextos del desarrollo: la especificación del sistema terminal en el embrión, el DV en el ovario, el patrón de venación del ala y el desarrollo de los genitales en los machos. Esta falta de función se da sin alterar la integridad de la proteína (Forés et al., 2017b; Fig. 2).

Dado que el motivo C1 tiene una función esencial para la capacidad represora de Cic, valoramos la posibilidad que fuera un motivo represor. Sin embargo, en un ensayo heterólogo ya habíamos visto que el motivo C1 no tiene una capacidad represora intrínseca como la tienen otros motivos de Cic como por ejemplo el motivo N2 (Forés et al., 2015; Fig. 4). Sorprendentemente, hemos visto que el C1 parece tener un comportamiento dependiente de la HMG-box, ya que cuando llevamos a Cic a reconocer genes diana diferentes a las suyos, sustituyendo la HMG-box por otro motivo de unión a ADN, el motivo C1 se hace dispensable (Forés et al., 2017b; Fig. 3). Estos resultados, nos llevaron a pensar que la función del motivo C1 podría estar relacionada con la HMG-box.

De manera similar como ocurre en *Drosophila*, en un ensayo de represión realizado en células humanas en cultivo, mutaciones del motivo C1 disminuyen la capacidad de CIC de reprimir sus dianas, tal como lo hacen mutaciones en la HMG-box. Además, en un

ensayo de inmunoprecipitación de cromatina, mutaciones tanto en la HMG-box como en el motivo C1 disminuyen la capacidad de CIC de unirse a las regiones promotoras de sus genes diana, indicando un requerimiento de ambos dominios para la asociación de CIC a estas regiones reguladoras (Forés et al., 2017b; Fig. 4).

3.2. El motivo C1 coopera con la HMG-box en la unión a ADN

Para clarificar la función del motivo C1 en la asociación de Cic a sus dianas, testamos la capacidad de Cic de unirse a ADN en presencia o ausencia del C1. Contrariamente a lo esperado, hemos visto que la HMG-box es insuficiente para mediar una asociación efectiva a ADN, de manera que tanto la HMG-box como el motivo C1 son necesarios para una correcta unión, tanto en *Drosophila* como en humanos. Las uniones detectadas son resultado de la acción de los dos dominios y son específicas, ya que mutaciones en cada uno de los motivos, así como en los lugares CBS eliminan la unión (Forés et al., 2017b; Fig. 5).

Estudiamos de qué manera el C1 puede estar actuando en la unión a ADN junto con la HMG-box. Se ha visto que los factores de transcripción con dominios HMG-box son capaces de reconocer secuencias específicas de ADN, pero la unión de la HMG-box sola a estos lugares no es suficiente para una selección apropiada de los genes diana y por lo tanto para ejercer una respuesta transcripcional apropiada (Revisado en Štros et al., 2007; Malarkey and Churchill, 2012; Kamachi and Kondoh, 2013). Por ejemplo, todas las proteínas Sox reconocen secuencias similares en el ADN (Badis et al., 2009; Kondoh and Kamachi, 2010) y expresándose más de una proteína Sox en un mismo momento y en un mismo tejido, cada una regula a sus propios genes diana. Esto parece indicar que hay algún otro elemento además de la HMG-box encargado de mediar la especificidad espacio-temporal. La selectividad de las proteínas Sox viene dada por su capacidad de formar complejos con otros factores de transcripción heterólogos o con otras proteínas Sox formando dímeros. De este modo, no sólo los lugares de unión a Sox son importantes para crear la especificidad y una unión efectiva a ADN, sino que lo son también lugares de unión adyacentes para otros factores. Las interacciones entre proteínas Sox con los diferentes factores se dan a través de un dominio de

dimerización cercano en la HMG-box o, a veces, a través de la propia HMG-box (Kamachi et al., 1999; Kamachi and Kondoh, 2013).

Estudiamos la posibilidad que la HMG-box o el C1 estuvieran mediando dimerización o reclutamiento de otros factores, realizando un mecanismo similar al que utilizan las proteínas Sox. Los resultados obtenidos indican que ni la HMG-box ni el C1 son necesarios para realizar un mecanismo similar. Una proteína compuesta únicamente por la HMG-box y el motivo C1 es activa en ausencia de cofactores y no media dimerización, de manera que ninguno de estos dos elementos tienen una función similar para mediar una unión efectiva al ADN (Forés et al., 2017b; Fig. 5). De todos modos, no podemos descartar la posibilidad de que otras regiones de la proteína (no testadas en este ensayo) puedan estar mediando este tipo de interacciones, aunque si así fuera, estas interacciones parecen ser dispensables para la función de la proteína *in vivo* ya que un transgén que contiene únicamente el N2, la HMG-box y el C1 puede rescatar la función de la proteína en un fondo mutante para *cic* (Forés et al., 2017b; Fig. S4).

3.3. El módulo HMG-C1 reconoce lugares octaméricos en el ADN

Hemos visto que Cic no se está uniendo a ADN como lo hacen las proteínas Sox, por lo que valoramos la posibilidad que lo estuviera haciendo como otras subfamilias HMG-box, como por ejemplo TCF. Algunas formas de TCF contienen un dominio de unión a ADN adicional a la HMG-box. Se trata de un dominio de dedos de zinc llamado *C-clamp*, el cual se caracteriza por unirse a lugares específicos cercanos al lugar de reconocimiento de la HMG-box (generalmente se une a menos de 10 nucleótidos de distancia del lugar de unión de la HMG-box). La presencia de ambos motivos de unión a ADN aumenta la fuerza y la especificidad de la unión de TCF a sus dianas (Atcha et al., 2007; Chang et al., 2008; Hoverter et al., 2012; Archbold et al., 2014; Bhambhani et al., 2014; Hoverter et al., 2014; Ravindranath and Cadigan, 2014).

Para estudiar la posibilidad que el C1 tuviese una función similar al motivo *C-Clamp* de TCF, hemos analizado las secuencias adyacentes a los CBSs entre diferentes genes diana, pero estos análisis no han revelado ninguna región conservada que pudiera ser un lugar de unión específico al motivo C1. Además, Cic se une con la misma afinidad y

de manera específica a sus genes diana independientemente de la secuencia que flanquee el octámero, incluso cuando la región colindante al CBS se ha sintetizado al azar (Forés et al., 2017b; Fig. 6).

Por otro lado, valoramos la posibilidad que las uniones en los ensayos de diferencia de movilidad electroforética (EMSA), se pudieran deber a un exceso de proteína, o por estar en unas condiciones óptimas de unión. En este escenario, las condiciones permitirían la unión de Cic a sus dianas independientemente de la conveniencia de las secuencias flanqueantes al CBS. Para descartar esta opción, diseñamos un ensayo *in vivo*. En este ensayo vimos que en el embrión de *Drosophila*, las secuencias octaméricas son suficientes para mediar represión a través de Cic, demostrando que el motivo C1 no actúa reconociendo secuencias específicas adyacentes al lugar de unión de la HMG-box como lo hace el motivo *C-clamp* (Forés et al., 2017b; Fig. 6).

Ante estos resultados, investigamos si el C1 se está uniendo de manera inespecífica al ADN en las regiones adyacentes al octámero. Hemos establecido que para unirse a ADN *in vitro* de una manera robusta, efectiva y específica, es suficiente con 18 pb, lo que parece sugerir que la HMG-box y el C1 se están uniendo a un núcleo relativamente invariable. De todos modos, no descartamos la posibilidad que tanto la HMG-box como el C1 estén contactando algún nucleótido fuera del octámero.

En general, los resultados obtenidos demuestran que Cic emplea un modo distinto de unirse a ADN al del resto de proteínas HMG-box. Mientras que las proteínas Sox necesitan asociarse con otros factores para aumentar la especificidad y la afinidad en la unión a ADN (Kamachi and Kondoh, 2013) y las proteínas TCF requieren de un dominio de unión a ADN específico de secuencia para aumentar la afinidad a ADN, Cic requiere del motivo C1. Este motivo se distingue de los dominios conservados en otras proteínas HMG-box en que está localizado lejos de la HMG-box, no se une a ADN por sí solo, no media dimerización ni oligomerización y no está implicado en el reconocimiento de secuencias específicas. Como las mutaciones del C1 no eliminan completamente la función de Cic *in vivo*, puede ser que el C1 actúe como potenciador de la unión a ADN. De hecho, los lugares de unión a Cic fueron identificados con una unión que implicaba únicamente la HMG-box (Kawamura-Saito et al., 2006), esto sugiere que la especificidad a los sitios la estaría dando la HMG-box y que el C1 estaría

potenciando la unión a estos lugares. Además, el C1 contiene muchos aminoácidos básicos conservados en su secuencia, los cuales podrían establecer contactos de baja afinidad con el ADN. De todos modos, no podemos descartar que el motivo C1 esté interaccionando con la HMG-box para modificar su plegamiento o estructura y facilitar un mejor acceso al ADN.

Dados los resultados, favorecemos un modelo de unión a ADN parecido al que utilizan las proteínas Pit-1 y Oct-1, las cuales tienen dominios POU de unión a ADN. La estructura del dominio POU es única entre los dominios de unión a ADN porque contiene dos subunidades separadas que cooperan funcionalmente como una unidad para la interacción con ADN. Las dos subunidades del dominio son estructuras distintas que han co-evolucionado y cooperan en la unión a ADN, de manera que ambos elementos son responsables de la especificidad de secuencia. Estos dos subelementos no interaccionan directamente entre ellos y se ha sugerido que la falta de interacción enfatiza la importancia de la secuencia que las separa: un *linker* flexible que permite a estos dos elementos funcionar de manera interdependiente como unidad (revisado en Herr and Cleary, 1995). De igual modo, la HMG-box y el motivo C1 no tienen similitud de secuencia y han co-evolucionado ya que ambos motivos se encuentran presentes en todas las proteínas Cic conocidas, de modo que nuestra hipótesis es que no son dos motivos independientes, sino que, tal como ocurre con los dominios POU, son una estructura bipartita de unión a ADN. La diferencia con las proteínas POU es la secuencia que separa las dos subunidades. No sabemos con exactitud qué tipo de estructura presenta la región que separa la HMG-box y el C1 en Cic, que en todos los casos analizados se trata de una secuencia relativamente larga y modelos de predicción de estructura indican que se trata de una zona desestrucurada. Además, hemos visto que separando la HMG-box y el motivo C1 tanto por un *linker* rígido como por un *linker* flexible no se afecta la unión a ADN. Sin embargo, la unión se pierde cuando se intercambia la posición de las dos subunidades, lo que podría indicar algún tipo de impedimento estérico en el acceso al ADN (Forés et al., 2017b; Fig. 5).

Por último, cabe destacar que para terminar de elucidar las bases moleculares de la función del C1, hemos iniciado un proyecto en colaboración para cristalizar la proteína HMG-box-C1 con el ADN. La cristalización resolvería los aspectos que se desconocen

de la función del motivo C1 como por ejemplo si interacciona con la HMG-box para cambiar su conformación y aumentar de algún modo su unión a ADN, o si contacta directamente al ADN.

3.4. El motivo C1 es indispensable para las funciones oncogénicas y supresoras tumorales de CIC

El patrón de mutaciones que afectan a CIC en oligodendroglioma y el patrón de las translocaciones que se dan en las quimeras CIC-DUX4 parecen indicar que el C1 es imprescindible para la función de CIC en los procesos donde actúa como represor o como activador de la transcripción. Hemos demostrado que algunas de las mutaciones recurrentes que se encuentran en oligodendroglioma afectan a la unión a ADN (Forés et al., 2017b; Fig. 1,4,5), por lo que el C1 es imprescindible en la función de la proteína como supresor tumoral.

Por otro lado, parece lógico pensar que, si el C1 está implicado en la unión de CIC a sus genes diana, una función oncogénica pasaría por el reconocimiento de estos genes y la unión a sus regiones reguladoras para ejercer la función activadora de la proteína. Investigamos si el motivo C1 es necesario para la activación y, efectivamente, hemos visto que el motivo C1 es imprescindible también para esta función oncogénica. Hemos mimetizado una forma activadora de Cic fusionándola con DUX4 y la hemos expresado en el disco imaginal de ala de la mosca. Esta forma lleva a la activación de los genes diana de Cic de forma dependiente del motivo C1, confirmando el papel del motivo en los dos contextos tumorales en los que Cic está implicado (Forés et al., 2017b; Fig. 7).

Nuestros resultados explicarían, por lo tanto, el patrón de mutaciones en oligodendroglioma, otros tipos de cáncer y metástasis que afectan los dominios HMG-box y C1. El modelo que proponemos es que estas mutaciones provocarían una unión insuficiente al ADN y la consiguiente desrepresión de los genes *ETV/PEA3*, los cuales codifican para los factores de transcripción ETS extensivamente implicados en tumorogénesis (Gleize et al., 2015; Padul et al., 2015; Okimoto et al., 2016). Por otro lado, en el caso de las quimeras CIC-DUX4, la presencia del motivo C1 favorecería la unión de la quimera activadora a los genes diana de CIC, resultando en una activación aberrante de los genes *ETV*.

Curiosamente, se han encontrado casos aislados de tumores de tipo *Ewing-Like* donde la translocación entre *CIC* y *DUX4* se da inmediatamente antes o en el *core* del motivo C1, yendo esto en contra de nuestros resultados. Sin embargo, la relevancia de estos casos viene dada por el hecho de que la translocación no se da en pauta, sino que altera el marco de lectura de la proteína DUX4, de manera que la quimera resultante *CIC-DUX4* codifica para una forma truncada de CIC en el C1 y no codifica para DUX4 (Italiano et al., 2012; Machado et al., 2013; Bielle et al., 2014; Tardío et al., 2015; Gambarotti et al., 2016). La explicación más sencilla para estos casos es que no se trata de una forma activadora de CIC sino una forma inactiva, de manera que esta quimera llevaría a la activación de los genes *ETV* por un mecanismo de desrepresión en lugar de una activación directa. Haría falta confirmar si en estos casos, tal como ocurre en oligodendroglioma cuando *CIC* está mutado, hay una pérdida de heterocigosidad debido a la translocación, de manera que se pierde una de las copias de *CIC* (Bettegowda et al., 2011), y verificar que no hay ninguna copia de la proteína que aún sea funcional.

Por último, las translocaciones *CIC-DUX4* no son las únicas que se han encontrado en sarcomas de tipo *Ewing-like* y que implican a *CIC*, sino que también se han encontrado translocaciones entre *CIC* y *FOXO4*. Solo se han encontrado 3 casos de sarcomas de tipo *Ewing-like* con translocaciones *CIC-FOXO4*. En uno de ellos la translocación se da después del C1 tal como ocurre con la mayoría de translocaciones *CIC-DUX4* (Brohl et al., 2014). En los dos casos restantes, la translocación se da en el mismo punto de la secuencia de *CIC*, inmediatamente antes del C1 (Solomon et al., 2014; Sugita et al., 2014). En estos casos, sin embargo, no se ha comprobado que esta translocación produzca una activación de los genes diana de CIC, y como en los casos de *CIC-DUX4* tampoco se ha comprobado si hay pérdida de heterocigosidad.

Los resultados de este trabajo suponen un avance en la comprensión del mecanismo de acción de Cic, lo que en un futuro facilitará la interpretación de mutaciones en este factor.

En conjunto, los resultados mostrados en esta tesis esclarecen aspectos clave del mecanismo de acción de Cic. Por un lado, se ha mostrado el mecanismo que Cic utiliza

en el embrión temprano de dípteros a través del correpresor Gro. Por otro lado, se ha aportado información acerca de los mecanismos de regulación de la actividad de Cic en respuesta a fosforilación. Por último, hemos caracterizado el modo que Cic utiliza para unirse a ADN en *Drosophila* y en humanos, además de explicar el patrón de mutaciones que afectan a CIC en diferentes tipos de cáncer. Nuestros resultados abren nuevos interrogantes acerca del mecanismo de Cic que aún se desconocen y que sería interesante seguir estudiando en un futuro, ya que entender el mecanismo que utiliza Cic para reprimir sus genes diana por debajo de las vías de señalización RTK ayudaría, entre otras muchas cosas, a la comprensión del mecanismo por el que la vía lleva producir múltiples enfermedades.

CONCLUSIONES

1. Cic ejerce actividades dependientes de Gro en el embrión temprano de *Drosophila*, e independientes en el disco imaginal de ala y las células foliculares.
2. El motivo N2 se encuentra únicamente en la forma Cic-S de *Drosophila* y media la interacción con Gro en el embrión temprano.
3. El motivo N2 y la isoforma Cic-S se originaron durante la aparición de los primeros dípteros hace 250 millones de años.
4. El motivo C2 es esencial para las funciones reguladoras de Cic por las vías RTK tanto en la forma Cic-L como en la forma Cic-S.
5. Cic está regulado negativamente por Mnb/Wap para controlar el crecimiento de diferentes tejidos así como para el establecimiento del patrón de vena en el ala.
6. La fosforilación de Cic por MAPK y Mnb/Wap tiene efectos aditivos en la regulación de la actividad de Cic en el disco imaginal de ala.
7. Cic se une a ADN a través de la HMG-box con la ayuda del motivo C1 el cual actúa como potenciador de la unión.
8. La unión de Cic a ADN utiliza un mecanismo distinto al del resto de proteínas HMG-box.
9. El motivo C1 es indispensable para las funciones de CIC como supresor tumoral y para las funciones oncogénicas de las quimeras CIC-DUX4 en humanos.

BIBLIOGRAFÍA

Ajuria, L., Nieva, C., Winkler, C., Kuo, D., Samper, N., Andreu, M. J., Helman, A., González-Crespo, S., Paroush, Z., Courey, A. J., et al. (2011). Capicua DNA-binding sites are general response elements for RTK signaling in *Drosophila*. *Development* **138**, 915-924.

Andreu, M. J., Ajuria, L., Samper, N., González-Pérez, E., Campuzano, S., González-Crespo, S. and Jiménez, G. (2012a). EGFR-dependent downregulation of Capicua and the establishment of *Drosophila* dorsoventral polarity. *Fly* **6**, 234-239.

Andreu, M. J., González-Pérez, E., Ajuria, L., Samper, N., González-Crespo, S., Campuzano, S. and Jiménez, G. (2012b). Mirror represses *pipe* expression in follicle cells to initiate dorsoventral axis formation in *Drosophila*. *Development* **139**, 1110-1114.

Archbold, H. C., Broussard, C., Chang, M. V. and Cadigan, K. M. (2014). Bipartite recognition of DNA by TCF/Pangolin is remarkably flexible and contributes to transcriptional responsiveness and tissue specificity of Wingless signaling. *PLoS Genet* **10**, e1004591.

Aronson, B. D., Fisher, A. L., Blechman, K., Caudy, M. and Gergen, J. P. (1997). Groucho-dependent and -independent repression activities of Runt domain proteins. *Mol Cell Biol* **17**, 5581-5587.

Astigarraga, S., Grossman, R., Díaz-Delfín, J., Caelles, C., Paroush, Z. and Jiménez, G. (2007). A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. *EMBO J* **26**, 668-677.

Atcha, F. A., Syed, A., Wu, B., Hoverter, N. P., Yokoyama, N. N., Ting, J. H. T., Munguia, J. E., Mangalam, H. J., Marsh, J. L. and Waterman, M. L. (2007). A unique DNA binding domain converts T-cell factors into strong Wnt effectors. *Mol Cell Biol* **27**, 8352-8363.

Atkey, M. R., Boisclair Lachance, J. F., Walczak, M., Rebello, T. and Nilson, L. A. (2006). Capicua regulates follicle cell fate in the *Drosophila* ovary through repression of *mirror*. *Development* **133**, 2115-2123.

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720-1723.

Berg, C. A. (2005). The *Drosophila* shell game: patterning genes and morphological change. *Trends Genet* **21**, 346-355.

Bettegowda, C., Agrawal, N., Jiao, Y., Sausen, M., Laura, D., Hruban, R. H., Rodriguez, F. J., Cahill, D. P., McLendon, R., Riggins, G., et al. (2011). Mutations in *CIC* and *FUBP1* contribute to human oligodendroglioma. *Science* **333**, 1453-1455.

Bhambhani, C., Ravindranath, A. J., Mentink, R. A., Chang, M. V., Betist, M. C., Yang, Y. X., Koushika, S. P., Korswagen, H. C. and Cadigan, K. M. (2014). Distinct DNA binding sites contribute to the TCF transcriptional switch in *C. elegans* and *Drosophila*. *PLoS Genet* **10**, e1004133.

Bielle, F., Zanello, M., Guillemot, D., Gil-Delgado, M., Bertrand, A., Boch, A.-L., Fréneaux, P. and Mokhtari, K. (2014). Unusual primary cerebral localization of a *CIC–DUX4* translocation tumor of the Ewing sarcoma family. *Acta Neuropathol* **128**, 309-311.

Blair, S. S. (2007). Wing vein patterning in *Drosophila* and the analysis of intracellular signaling. *Annu Rev Cell Dev Biol* **23**, 293-319.

Blasco, R. B., Francoz, S., Santamaría, D., Cañamero, M., Dubus, P., Charron, J., Baccarini, M. and Barbacid, M. (2011). c-Raf, but not B-Raf, is essential for development of K-Ras oncogene driven non-small cell lung carcinoma. *Cancer cell* **19**, 652-663.

Boisclair Lachance, J. F., Fregoso Lomas, M., Eleiche, A., Bouchard Kerr, P. and Nilson, L. A. (2009). Graded Egfr activity patterns *Drosophila* eggshell independently of autocrine feedback. *Development* **136**, 2893-2902.

Bowman, A. B., Lam, Y. C., Jafar-Nejad, P., Chen, H.-K., Richman, R., Samaco, R. C., Fryer, J. D., Kahle, J. J., Orr, H. T. and Zoghbi, H. Y. (2007). Duplication of *Atxn1L* suppresses SCA1 neuropathology by decreasing incorporation of polyglutamine-expanded ataxin-1 into native complexes. *Nat Genet* **39**, 373-379.

Brohl, A. S., Solomon, D. A., Chang, W., Wang, J., Song, Y., Sindiri, S., Patidar, R., Hurd, L., Chen, L., Shern, J. F., et al. (2014). The genomic landscape of the ewing sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet* **10**, e1004475.

Brönner, G. and Jäckle, H. (1996). Regulation and function of the terminal gap gene *huckebein* in the *Drosophila* blastoderm. *Int J Dev Biol* **40**, 157-165.

Brunner, D., Ducker, K., Oellers, N., Hafen, E., Scholzi, H. and Klambt, C. (1994). The ETS domain protein Pointed-P2 is a target of MAP kinase in the Sevenless signal transduction pathway. *Nature* **370**, 386-389.

Canon, J. and Banerjee, U. (2003). In vivo analysis of a developmental circuit for direct transcriptional activation and repression in the same cell by a Runx protein. *Genes Dev* **17**, 838-843.

Casali, A. and Casanova, J. (2001). The spatial control of Torso RTK activation: a C-terminal fragment of the Trunk protein acts as a signal for Torso receptor in the *Drosophila* embryo. *Development* **128**, 1709-1715.

Casanova, J. (1991). Interaction between *torso* and *dorsal*, two elements of different transduction pathways in the *Drosophila* embryo. *Mech Dev* **36**, 41-45.

Casanova, J., Furriols, M., McCormick, C. A. and Struhl, G. (1995). Similarities between *trunk* and *spätzle*, putative extracellular ligands specifying body pattern in *Drosophila*. *Genes Dev* **9**, 2539-2544.

Chan, A. K. Y., Pang, J. C.-S., Chung, N. Y.-F., Li, K. K. W., Poon, W. S., Chan, D. T. M., Shi, Z., Chen, L., Zhou, L. and Ng, H.-K. (2014). Loss of CIC and FUBP1 expressions are potential markers of shorter time to recurrence in oligodendroglial tumors. *Mod Pathol* **27**, 332-342.

Chang, M. V., Chang, J. L., Gangopadhyay, A., Shearer, A. and Cadigan, K. M. (2008). Activation of Wingless targets requires bipartite recognition of DNA by TCF. *Curr Biol* **18**, 1877-1881.

Chen, G. and Courey, A. J. (2000). Groucho/TLE family proteins and transcriptional repression. *Gene* **249**, 1-16.

Chen, G., Fernandez, J., Mische, S. and Courey, A. J. (1999). A functional interaction between the histone deacetylase Rpd3 and the corepressor Groucho in *Drosophila* development. *Genes Dev* **13**, 2218-2230.

Cheung, L. S., Schüpbach, T. and Shvartsman, S. Y. (2011). Pattern formation by receptor tyrosine kinases: analysis of the Gurken gradient in *Drosophila* oogenesis. *Curr Opin Genet Dev* **21**, 719-725.

Chittaranjan, S., Chan, S., Yang, C., Yang, K. C., Chen, V., Moradian, A., Firme, M., Song, J., Go, N. E., Blough, M. D., et al. (2014). Mutations in CIC and IDH1 cooperatively regulate 2-hydroxyglutarate levels and cell clonogenicity. *Oncotarget* **5**, 7960-7979.

Choi, E. Y. K., Thomas, D. G., McHugh, J. B., Patel, R. M., Roulston, D., Schuetze, S. M., Chugh, R., Biermann, J. S. and Lucas, D. R. (2013). Undifferentiated small round cell sarcoma with t(4;19)(q35;q13.1) *CIC-DUX4* fusion: a novel highly aggressive soft tissue tumor with distinctive histopathology. *Am J Surg Pathol* **37**, 1379-1386.

Cinnamon, E., Helman, A., Ben-Haroush Schyr, R., Orian, A., Jiménez, G. and Paroush, Z. (2008). Multiple RTK pathways downregulate Groucho-mediated repression in *Drosophila* embryogenesis. *Development* **135**, 829-837.

Copley, R. R. (2005). The EH1 motif in metazoan transcription factors. *BMC Genomics* **6**, 169.

Crespo-Barreto, J., Fryer, J. D., Shaw, C. A., Orr, H. T. and Zoghbi, H. Y. (2010). Partial loss of Ataxin-1 function contributes to transcriptional dysregulation in spinocerebellar ataxia type 1 pathogenesis. *PLoS Genet* **6**, e1001021.

de Celis, J. F., Bray, S. and Garcia-Bellido, A. (1997). Notch signalling regulates *veinlet* expression and establishes boundaries between veins and interveins in the *Drosophila* wing. *Development* **124**, 1919-1928.

Degoutin, J. L., Milton, C. C., Yu, E., Tipping, M., Bosveld, F., Yang, L., Bellaiche, Y., Veraksa, A. and Harvey, K. F. (2013). Riquiqui and Minibrain are regulators of the Hippo pathway downstream of Dachsous. *Nat Cell Biol* **15**, 1176-1185.

Delidakis, C., Preiss, A., Hartley, D. A. and Artavanis-Tsakonas, S. (1991). Two genetically and molecularly distinct functions involved in early neurogenesis reside within the Enhancer of split locus of *Drosophila melanogaster*. *Genetics* **129**, 803-823.

Diaz-Benjumea, F. J. and Hafen, E. (1994). The *sevenless* signalling cassette mediates *Drosophila* EGF receptor function during epidermal development. *Development* **120**, 569-578.

Dissanayake, K., Toth, R., Blakey, J., Olsson, O., Campbell, D. G., Prescott, A. R. and MacKintosh, C. (2011). ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription factors via the transcriptional repressor *capicúa*. *Biochem J* **433**, 515-525.

Dorman, J. B., James, K. E., Fraser, S. E., Kiehart, D. P. and Berg, C. A. (2004). *bullwinkle* is required for epithelial morphogenesis during *Drosophila* oogenesis. *Dev Biol* **267**, 320-341.

Dubnicoff, T., Valentine, S. A., Chen, G., Shi, T., Lengyel, J. A., Paroush, Z. and Courey, A. J. (1997). Conversion of Dorsal from an activator to a repressor by the global corepressor Groucho. *Genes Dev* **11**, 2952-2957.

Duffy, J. B. and Perrimon, N. (1994). The Torso pathway in *Drosophila*: lessons on Receptor Tyrosine Kinase signaling and pattern formation. *Dev Biol* **166**, 380-395.

El-Sherif, E., Averof, M. and Brown, S. J. (2012). A segmentation clock operating in blastoderm and germband stages of *Tribolium* development. *Development* **139**, 4341-4346.

Fisher, A. L., Ohsako, S. and Caudy, M. (1996). The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain. *Mol Cell Biol* **16**, 2670-2677.

Forés, M., Ajuria, L., Samper, N., Astigarraga, S., Nieva, C., Grossman, R., González-Crespo, S., Paroush, Z. and Jiménez, G. (2015). Origins of context-dependent gene repression by Capicua. *PLoS Genet* **11**, e1004902.

Forés, M., Papagianni, A., Rodríguez-Muñoz, L. and Jiménez, G. (2017a). Using CRISPR-Cas9 to Study ERK Signaling in *Drosophila*. In *ERK Signaling: Methods and Protocols* (ed. G. Jimenez), pp. 353-365. New York, NY: Springer New York.

Forés, M., Simón-Carrasco, L., Ajuria, L., Samper, N., González-Crespo, S., Drosten, M., Barbacid, M. and Jiménez, G. (2017b). A new mode of DNA binding distinguishes Capicua from other HMG-box factors and explains its mutation patterns in cancer. *PLoS Genet* **13**, e1006622.

Freeman, M., Klämbt, C., Goodman, C. S. and Rubin, G. M. (1992). The *argos* gene encodes a diffusible factor that regulates cell fate decisions in the *Drosophila* eye. *Cell* **69**, 963-975.

Fryer, J. D., Yu, P., Kang, H., Mandel-Brehm, C., Carter, A. N., Crespo-Barreto, J., Gao, Y., Flora, A., Shaw, C., Orr, H. T., et al. (2011). Exercise and genetic rescue of SCA1 via the transcriptional repressor Capicua. *Science* **334**, 690-693.

Fuchs, A., Cheung, L. S., Charbonnier, E., Shvartsman, S. Y. and Pyrowolakis, G. (2012). Transcriptional interpretation of the EGF receptor signaling gradient. *Proc Natl Acad Sci USA* **109**, 1572-1577.

Futran, A. S., Kyin, S., Shvartsman, S. Y. and Link, A. J. (2015). Mapping the binding interface of ERK and transcriptional repressor Capicua using photocrosslinking. *Proc Natl Acad Sci USA* **112**, 8590-8595.

Galant, R. and Carroll, S. B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* **415**, 910-913.

Gambarotti, M., Benini, S., Gamberi, G., Cocchi, S., Palmerini, E., Sbaraglia, M., Donati, D., Picci, P., Vanel, D., Ferrari, S., et al. (2016). *CIC–DUX4* fusion-positive round-cell sarcomas of soft tissue and bone: a single-institution morphological and molecular analysis of seven cases. *Histopathology* **69**, 624-634.

Garcia-Bellido, A. and de Celis, J. F. (1992). Developmental genetics of the venation pattern of *Drosophila*. *Annu Rev Genet* **26**, 277-304.

Gasperowicz, M. and Otto, F. (2005). Mammalian Groucho homologs: redundancy or specificity? *J Cell Biochem* **95**, 670-687.

Gilbert, S. F. (2003). *Developmental Biology*: Sunderland: Synauer Associates Inc.

Gleize, V., Alentorn, A., Connen de Kérillis, L., Labussière, M., Nadaradjane, A. A., Mundwiller, E., Ottolenghi, C., Mangesius, S., Rahimian, A., Ducray, F., et al. (2015). *CIC* inactivating mutations identify aggressive subset of 1p19q codeleted gliomas. *Ann Neurol* **78**, 355-374.

Goff, D. J., Nilson, L. A. and Morisato, D. (2001). Establishment of dorsal-ventral polarity of the *Drosophila* egg requires *capicua* action in ovarian follicle cells. *Development* **128**, 4553-4562.

Goldstein, R. E., Jimenez, G., Cook, O., Gur, D. and Paroush, Z. (1999). Huckebein repressor activity in *Drosophila* terminal patterning is mediated by Groucho. *Development* **126**, 3747-3755.

Graham, C., Chilton-MacNeill, S., Zielenska, M. and Somers, G. R. (2012). The *CIC-DUX4* fusion transcript is present in a subgroup of pediatric primitive round cell sarcomas. *Hum Pathol* **43**, 180-189.

Helman, A., Cinnamon, E., Mezuman, S., Hayouka, Z., Von Ohlen, T., Orian, A., Jiménez, G. and Paroush, Z. (2011). Phosphorylation of Groucho mediates RTK feedback inhibition and prolonged pathway target gene expression. *Curr Biol* **21**, 1102-1110.

Herr, W. and Cleary, M. A. (1995). The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev* **9**, 1679-1693.

Hittinger, C. T. and Carroll, S. B. (2008). Evolution of an insect-specific GROUCHO-interaction motif in the ENGRAILED selector protein. *Evol Dev* **10**, 537-545.

Hoverter, N. P., Ting, J. H., Sundaresh, S., Baldi, P. and Waterman, M. L. (2012). A WNT/p21 circuit directed by the C-clamp, a sequence-specific DNA binding domain in TCFs. *Mol Cell Biol* **32**, 3648-3662.

Hoverter, N. P., Zeller, M. D., McQuade, M. M., Garibaldi, A., Busch, A., Selwan, E. M., Hertel, K. J., Baldi, P. and Waterman, M. L. (2014). The TCF C-clamp DNA binding domain expands the Wnt transcriptome via alternative target recognition. *Nucleic Acids Res* **42**, 13615-13632.

Ip, T. Y., Kraut, R., Levine, M. and Rushlow, C. A. (1991). The *dorsal* morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in *Drosophila*. *Cell* **64**, 439-446.

Italiano, A., Sung, Y. S., Zhang, L., Singer, S., Maki, R. G. and Coindre, J. M. (2012). High prevalence of *CIC* fusion with Double-Homeobox (DUX4) transcription factors in *EWSR1*-negative undifferentiated small blue round cell sarcomas. *Genes Chromosomes Cancer* **51**, 207-218.

Jennings, B. H., Pickles, L. M., Wainwright, S. M., Roe, S. M., Pearl, L. H. and Ish-Horowicz, D. (2006). Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol Cell* **22**, 645-655.

Jennings, B. H., Wainwright, S. M. and Ish-Horowicz, D. (2008). Differential in vivo requirements for oligomerization during Groucho-mediated repression. *EMBO Rep* **9**, 76-83.

Jiang, H., Grenley, M. O., Bravo, M.-J., Blumhagen, R. Z. and Edgar, B. A. (2011). EGFR/Ras/MAPK signaling mediates adult midgut epithelial homeostasis and regeneration in *Drosophila*. *Cell Stem Cell* **8**, 84-95.

Jiang, J., Kosman, D., Ip, Y. T. and Levine, M. (1991). The *dorsal* morphogen gradient regulates the mesoderm determinant twist in early *Drosophila* embryos. *Genes Dev* **5**, 1881-1891.

Jiao, Y., Killela, P. J., Reitman, Z. J., Rasheed, B. A., Heaphy, C. M., Wilde, R. F. D., Rodriguez, F. J., Oba-shinjo, S. M., Kazue, S., Marie, N., et al. (2012). Frequent *ATRX*, *CIC*, *FUBP1* and *IDH1* mutations refine the classification of malignant gliomas. *Oncotarget* **3**, 709-722.

Jiménez, G., Guichet, A., Ephrussi, A. and Casanova, J. (2000). Relief of gene repression by Torso RTK signaling : role of *capicua* in *Drosophila* terminal and dorsoventral patterning. *Genes Dev* **14**, 224-231.

Jiménez, G., Paroush, Z. and Ish-Horowicz, D. (1997). Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev* **11**, 3072-3082.

Jiménez, G., Shvartsman, S. Y. and Paroush, Z. (2012). The Capicua repressor-a general sensor of RTK signaling in development and disease. *J Cell Sci* **125**, 1383-1391.

Jiménez, G., Verrijzer, C. P. and Ish-Horowicz, D. (1999). A conserved motif in Goosecoid mediates groucho-dependent repression in *Drosophila* embryos. *Mol Cell Biol* **19**.

Jin, Y., Ha, N., Forés, M., Xiang, J., Gläßer, C., Maldera, J., Jiménez, G. and Edgar, B. A. (2015). EGFR/Ras signaling controls *Drosophila* intestinal stem cell proliferation via Capicua-regulated genes. *PLoS Genet* **11**, e1005634.

Johnson, T. K., Henstridge, M. A., Herr, A., Moore, K. A., Whisstock, J. C. and Warr, C. G. (2015). Torso-like mediates extracellular accumulation of Furin-cleaved Trunk to pattern the *Drosophila* embryo termini. *Nat Commun* **6**, 8759.

Jordan, K. C., Clegg, N. J., Blasi, J. A., Morimoto, A. M., Sen, J., Stein, D., McNeill, H., Deng, W.-M., Tworoger, M. and Ruohola-Baker, H. (2000). The homeobox gene *mirror* links EGF signalling to embryonic dorso-ventral axis formation through Notch activation. *Nat Genet* **24**, 429-433.

Jurgens, G. and Hartenstein, V. (1993). The terminal regions of the body pattern. In *The development of Drosophila melanogaster* (ed. C. S. H. Laboratory), pp. 687-746.

Kamachi, Y., Cheah, K. S. E. and Kondoh, H. (1999). Mechanism of regulatory target selection by the SOX High-Mobility-Group domain proteins as revealed by comparison of SOX1/2/3 and SOX9. *Mol Cell Biol* **19**, 107-120.

Kamachi, Y. and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129-4144.

Kawamura-Saito, M., Yamazaki, Y., Kaneko, K., Kawaguchi, N., Kanda, H., Mukai, H., Gotoh, T., Motoi, T., Fukayama, M., Aburatani, H., et al. (2006). Fusion between *CIC* and *DUX4* up-regulates *PEA3* family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. *Hum Mol Genet* **15**, 2125-2137.

Kazemian, M., Blatti, C., Richards, A., McCutchan, M., Wakabayashi-Ito, N., Hammonds, A. S., Celniker, S. E., Kumar, S., Wolfe, S. A., Brodsky, M. H., et al. (2010). Quantitative analysis of the *Drosophila* segmentation regulatory network using pattern generating potentials. *PLoS Biol* **8**, e1000456.

Kelly, P. A. and Rahmani, Z. (2005). DYRK1A enhances the Mitogen-Activated Protein Kinase cascade in PC12 cells by forming a complex with Ras, B-Raf, and MEK1. *Mol Biol Cell* **16**, 3562-3573.

Kim, E., Lu, H.-C., Zoghbi, H. Y. and Song, J.-J. (2013). Structural basis of protein complex formation and reconfiguration by polyglutamine disease protein Ataxin-1 and Capicua. *Genes Dev* **27**, 590-595.

Kim, E., Park, S., Choi, N., Lee, J., Yoe, J., Kim, S., Jung, H. Y., Kim, K. T., Kang, H., Fryer, J. D., et al. (2015). Deficiency of Capicua disrupts bile acid homeostasis. *Sci Rep* **5**, 8272.

King, R. C. (1970). *Ovarian Development in Drosophila melanogaster*: New York: Academic Press.

Kirov, N., Zhelnin, L., Shah, J. and Rushlow, C. (1993). Conversion of a silencer into an enhancer: evidence for a co-repressor in dorsal-mediated repression in *Drosophila*. *EMBO J* **12**, 3193-3199.

Kondoh, H. and Kamachi, Y. (2010). SOX–partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *Int J Biochem Cell Biol* **42**, 391-399.

Krivy, K., Bradley-Gill, M.-R. and Moon, N.-S. (2013). Capicua regulates proliferation and survival of RB-deficient cells in *Drosophila*. *Biol Open* **2**, 183-190.

Lam, Y. C., Bowman, A. B., Jafar-Nejad, P., Lim, J., Richman, R., Fryer, J. D., Hyun, E. D., Duvick, L. a., Orr, H. T., Botas, J., et al. (2006). ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell* **127**, 1335-1347.

Lee, C.-J., Chan, W.-I., Cheung, M., Cheng, Y.-C., Appleby, V. J., Orme, A. T. and Scotting, P. J. (2002). CIC, a member of a novel subfamily of the HMG-box superfamily, is transiently expressed in developing granule neurons. *Mol Brain Res* **106**, 151-156.

Lee, Y., Fryer, J. D., Kang, H., Crespo-Barreto, J., Bowman, A. B., Gao, Y., Kahle, J. J., Hong, J. S., Kheradmand, F., Orr, H. T., et al. (2011). ATXN1 protein family and CIC regulate extracellular matrix remodeling and lung alveolarization. *Dev Cell* **21**, 746-757.

Li, D. and Roberts, R. (2001). WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases *Cell Mol Life Sci* **58**, 2085-2097.

Liao, S., Davoli, T., Leng, Y., Li, M. Z., Xu, Q. and Elledge, S. J. (2017). A genetic interaction analysis identifies cancer drivers that modify EGFR dependency. *Genes Dev* **31**, 184-196.

Lim, J., Crespo-Barreto, J., Jafar-Nejad, P., Bowman, A. B., Richman, R., Hill, D. E., Orr, H. T. and Zoghbi, H. Y. (2008). Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. *Nature* **452**, 713-718.

Löhr, U., Chung, H. R., Beller, M. and Jäckle, H. (2009). Antagonistic action of Bicoid and the repressor Capicua determines the spatial limits of *Drosophila* head gene expression domains. *Proc Natl Acad Sci USA* **106**, 21695-21700.

Lynch, J. A., Olesnicky, E. C. and Desplan, C. (2006). Regulation and function of *tailless* in the long germ wasp *Nasonia vitripennis*. *Dev Genes Evol* **216**, 493-498.

Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., et al. (2004). Activating mutations in the Epidermal Growth Factor Receptor underlying responsiveness of non–small-cell lung cancer to gefitinib. *N Engl J Med* **350**, 2129-2139.

Machado, I., Cruz, J., Lavernia, J., Rubio, L., Campos, J., Barrios, M., Grison, C., Chene, V., Pierron, G., Delattre, O., et al. (2013). Superficial *EWSR1*-negative undifferentiated small round cell sarcoma with CIC/*DUX4* gene fusion: a new variant of Ewing-like tumors with locoregional lymph node metastasis. *Virchows Arch* **463**, 837-842.

Malarkey, C. S. and Churchill, M. E. A. (2012). The high mobility group box: the ultimate utility player of a cell. *Trends Biochem Sci* **37**, 553-562.

Martin, J.-R., Raibaud, A. and Ollo, R. (1994). Terminal pattern elements in *Drosophila* embryo induced by the torso-like protein. *Nature* **367**, 741-745.

Mineo, A., Furriols, M. and Casanova, J. (2015). Accumulation of the *Drosophila* Torso-like protein at the blastoderm plasma membrane suggests that it translocates from the eggshell. *Development* **142**, 1299-1304.

Mok , T. S., Wu , Y.-L., Thongprasert , S., Yang , C.-H., Chu , D.-T., Saijo , N., Sunpaweravong , P., Han , B., Margono , B., Ichinose , Y., et al. (2009). Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361**, 947-957.

Neuman-Silberberg, F. S. and Schupbach, T. (1994). Dorsoventral axis formation in *Drosophila* depends on the correct dosage of the gene *gurken*. *Development* **120**, 2457.

Okimoto, R. A., Breitenbuecher, F., Olivas, V. R., Wu, W., Gini, B., Hofree, M., Asthana, S., Hrustanovic, G., Flanagan, J., Tulpule, A., et al. (2016). Inactivation of Capicua drives cancer metastasis. *Nat Genet* **49**, 87-96.

Orr, H. T., Chung, M.-y., Banfi, S., Kwiatkowski Jr., T. J., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P. W. and Zoghbi, H. Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* **4**, 221-226.

Padul, V., Epari, S., Moiyadi, A., Shetty, P. and Shirsat, N. V. (2015). ETV / Pea3 family transcription factor-encoding genes are overexpressed in *CIC*-mutant oligodendrogliomas. *Genes Chromosomes Cancer* **54**, 725-733.

Paez, J. G., Jänne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., Boggon, T. J., et al. (2004). *EGFR* mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497-1500.

Parkhurst, S. M. (1998). Groucho: making its Marx as a transcriptional co-repressor. *Trends Genet* **14**, 130-132.

Paroush, Z., Finley, R. L., Jr., Kidd, T., Wainwright, S. M., Ingham, P. W., Brent, R. and Ish-Horowicz, D. (1994). Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**, 805-815.

Paroush, Z., Wainwright, S. M. and Ish-Horowicz, D. (1997). Torso signalling regulates terminal patterning in *Drosophila* by antagonising Groucho-mediated repression. *Development* **124**, 3827-3834.

Pickles, L. M., Roe, S. M., Hemingway, E. J., Stifani, S. and Pearl, L. H. (2002). Crystal structure of the C-Terminal WD40 repeat domain of the human Groucho/TLE1 transcriptional corepressor. *Structure* **10**, 751-761.

Pignoni, F., Baldarelli, R. M., Steingrimsson, E., Diaz, R. J., Patapoutian, A., Merriam, J. R. and Lengyel, J. A. (1990). The *Drosophila* gene *tailless* is expressed at the embryonic termini and is a member of the steroid receptor superfamily. *Cell* **62**, 151-163.

Pozo, N., Zahonero, C., Fern, xE, ndez, P., Li, xF, ares, J. M., Ayuso, A., Hagiwara, M., et al. (2013). Inhibition of DYRK1A destabilizes EGFR and reduces EGFR-dependent glioblastoma growth. *J Clin Invest* **123**, 2475-2487.

Pridöhl, F., Weißkopf, M., Koniszewski, N., Sulzmaier, A., Uebe, S., Ekici, A. B. and Schoppmeier, M. (2017). Transcriptome sequencing reveals *maelstrom* as a novel target gene of the terminal-system in the red flour beetle *Tribolium castaneum*. *Development*.

Queenan, A. M., Ghabrial, A. and Schupbach, T. (1997). Ectopic activation of *torpedo/Egfr*, a *Drosophila* receptor tyrosine kinase, dorsalizes both the eggshell and the embryo. *Development* **124**, 3871-3880.

Ravindranath, A. and Cadigan, K. M. (2014). Structure-function analysis of the C-clamp of TCF/Pangolin in Wnt/ß-catenin signaling. *PLoS ONE* **9**, e86180.

Rebay, I. and Rubin, G. M. (1995). Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. *Cell* **81**, 857-866.

Reményi, A., Lins, K., Nissen, L. J., Reinbold, R., Schöler, H. R. and Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17**, 2048-2059.

Rittenhouse, K. R. and Berg, C. A. (1995). Mutations in the *Drosophila* gene *bullwinkle* cause the formation of abnormal eggshell structures and bicaudal embryos. *Development* **121**, 3023-3033.

Roch, F., Jiménez, G. and Casanova, J. (2002). EGFR signalling inhibits Capicua-dependent repression during specification of *Drosophila* wing veins. *Development* **129**, 993-1002.

Roth, S., Stein, D. and Nüsslein-Volhard, C. (1989). A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell* **59**, 1189-1202.

Rusch, J. and Levine, M. (1994). Regulation of the *dorsal* morphogen by the Toll and torso signaling pathways: a receptor tyrosine kinase selectively masks transcriptional repression. *Genes Dev* **8**, 1247-1257.

Rusch, J. and Levine, M. (1996). Threshold responses to the dorsal regulatory gradient and the subdivision of primary tissue territories in the *Drosophila* embryo. *Curr Opin Genet Dev* **6**, 416-423.

Rushlow, C. A., Han, K., Manley, J. L. and Levine, M. (1989). The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in *Drosophila*. *Cell* **59**, 1165-1177.

Sahm, F., Koelsche, C., Meyer, J., Pusch, S., Lindenberg, K., Mueller, W., Herold-Mende, C., von Deimling, A. and Hartmann, C. (2012). *CIC* and *FUBP1* mutations in oligodendrogliomas, oligoastrocytomas and astrocytomas. *Acta Neuropathol* **123**, 853-860.

Sanson, B. (2001). Generating patterns from fields of cells. Examples from *Drosophila* segmentation. *EMBO Rep* **2**, 1083-1088.

Schoppmeier, M. and Schröder, R. (2005). Maternal Torso signaling controls body axis elongation in a short germ insect. *Curr Biol* **15**, 2131-2136.

Schröder, R., Beermann, A., Wittkopp, N. and Lutz, R. (2008). From development to biodiversity—*Tribolium castaneum*, an insect model organism for short germband development. *Dev Genes Evol* **218**, 119-126.

Schröder, R., Eckert, C., Wolff, C. and Tautz, D. (2000). Conserved and divergent aspects of terminal patterning in the beetle *Tribolium castaneum*. *Proc Natl Acad Sci USA* **97**, 6591-6596.

Schüpbach, T. (1987). Germ line and soma cooperate during oogenesis to establish the dorsoventral pattern of egg shell and embryo in *Drosophila melanogaster*. *Cell* **49**, 699-707.

Schüpbach, T. and Wieschaus, E. (1986). Germline autonomy of maternal-effect mutations altering the embryonic body pattern of *Drosophila*. *Dev Biol* **113**, 443-448.

Sen, J., Goltz, J. S., Stevens, L. and Stein, D. (1998). Spatially restricted expression of *pipe* in the *Drosophila* egg chamber defines embryonic dorsal-ventral polarity. *Cell* **95**, 471-481.

Seshagiri, S., Stawiski, E. W., Durinck, S., Modrusan, Z., Storm, E. E., Conboy, C. B., Chaudhuri, S., Guan, Y., Janakiraman, V., Jaiswal, B. S., et al. (2012). Recurrent R-spondin fusions in colon cancer. *Nature* **488**, 660-664.

Sharifnia, T., Rusu, V., Piccioni, F., Bagul, M., Imielinski, M., Cherniack, A. D., Pedamallu, C. S., Wong, B., Wilson, F. H., Garraway, L. A., et al. (2014). Genetic modifiers of EGFR dependence in non-small cell lung cancer. *Proc Natl Acad Sci USA* **111**, 18661-18666.

Shilo, B.-Z. (2003). Signaling by the *Drosophila* epidermal growth factor receptor pathway during development. *Exp Cell Res* **284**, 140-149.

Shilo, B.-Z. (2005). Regulating the dynamics of EGF receptor signaling in space and time. *Development* **132**, 4017-4027.

Sjöblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., et al. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-274.

Skurat, A. V. and Dietrich, A. D. (2004). Phosphorylation of Ser640 in muscle glycogen synthase by DYRK family protein kinases. *J Biol Chem* **279**, 2490-2498.

Smith, S. T. and Jaynes, J. B. (1996). A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo. *Development* **122**, 3141-3150.

Solomon, D. A., Brohl, A. S., Khan, J. and Miettinen, M. (2014). Clinicopathologic features of a second patient with Ewing-like sarcoma harboring CIC-FOXO4 gene fusion. *Am J Surg Pathol* **38**, 1724-1725.

Spradling, A. C. (1993). Developmental genetics of oogenesis. In *Developmental genetics of oogenesis* (ed. M.-A. B), pp. 1-70: Cold Spring Harbor Laboratory.

St Johnston, D. and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201-219.

Steingrimsson, E., Pignoni, F., Liaw, G. and Lengyel, J. (1991). Dual role of the *Drosophila* pattern gene *tailless* in embryonic termini. *Science* **254**, 418-421.

Stevens, L. M., Beuchle, D., Jurcsak, J., Tong, X. and Stein, D. (2003). The *Drosophila* embryonic patterning determinant Torsolike is a component of the eggshell. *Curr Biol* **13**, 1058-1063.

Stevens, L. M., Frohnhofer, H. G., Klingler, M. and Nusslein-Volhard, C. (1990). Localized requirement for *torso-like* expression in follicle cells for development of terminal anlagen of the *Drosophila* embryo. *Nature* **346**, 660-663.

Steward, R. (1989). Relocalization of the dorsal protein from the cytoplasm to the nucleus correlates with its function. *Cell* **59**, 1179-1188.

Stewart, E. L., Tan, S. Z., Liu, G. and Tsao, M.-S. (2015). Known and putative mechanisms of resistance to EGFR targeted therapies in NSCLC patients with EGFR mutations—a review. *Trans Lung Cancer Res.* **4**, 67-81.

Štros, M., Launholt, D. and Grasser, K. D. (2007). The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell Mol Life Sci* **64**, 2590-2606.

Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev* **12**, 599-606.

Sugita, S., Arai, Y., Tonooka, A., Hama, N., Totoki, Y., Fujii, T., Aoyama, T., Asanuma, H., Tsukahara, T., Kya, M., et al. (2014). A novel *CIC-FOXO4* gene fusion in undifferentiated small round cell sarcoma, a genetically distinct variant of Ewing-like sarcoma. *Am J Surg Pathol* **38**, 1571-1576.

Tardío, J. C., Machado, I., Navarro, L., Idrovo, F., Sanz-Ortega, J., Pellín, A. and Llombart-Bosch, A. (2015). Ewing-like sarcoma with *CIC-DUX4* gene fusion in a patient with neurofibromatosis type 1. A hitherto unreported association. *Pathol Res Pract.* **211**, 877-882.

Technau, M., Knispel, M. and Roth, S. (2012). Molecular mechanisms of EGF signaling-dependent regulation of *pipe,* a gene crucial for dorsoventral axis formation in *Drosophila*. *Dev Genes Evol* **222**, 1-17.

Tejedor, F., Zhu, X. R., Kaltenbach, E., Ackermann, A., Baumann, A., Canal, I., Heisenberg, M., Fischbach, K. F. and Pongs, O. (1995). *minibrain*: A new protein kinase family involved in postembryonic neurogenesis in *Drosophila*. *Neuron* **14**, 287-301.

Tolkunova, E. N., Fujioka, M., Kobayashi, M., Deka, D. and Jaynes, J. B. (1998). Two distinct types of repression domain in Engrailed: one interacts with the Groucho corepressor and is preferentially active on integrated target genes. *Mol Cell Biol* **18**, 2804-2814.

Tootle, T. L. and Rebay, I. (2005). Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays* **27**, 285-298.

Tseng, A. S. K., Tapon, N., Kanda, H., Cigizoglu, S., Edelmann, L., Pellock, B., White, K. and Hariharan, I. K. (2007). Capicua regulates cell proliferation downstream of the Receptor Tyrosine Kinase/Ras signaling pathway. *Curr Biol* **8**, 728-733.

Tsuda, H., Jafar-Nejad, H., Patel, A. J., Sun, Y., Chen, H.-K., Rose, M. F., Venken, K. J. T., Botas, J., Orr, H. T., Bellen, H. J., et al. (2005). The AXH domain of Ataxin-1 mediates neurodegeneration through its interaction with Gfi-1/Senseless proteins. *Cell* **122**, 633-644.

Turki-Judeh, W. and Courey, A. J. (2012). Groucho: A Corepressor with Instructive Roles in Development. In *Current Topics in Developmental Biology* (ed. P. Serge & P. François), pp. 65-96: Academic Press.

Ullrich, A. and Schlessinger, J. (1990). Signal transduction by receptors with tyrosine kinase activity. *Cell* **61**, 203-212.

Valentine, S. A., Chen, G., Shandala, T., Fernandez, J., Mische, S., Saint, R. and Courey, A. J. (1998). Dorsal-mediated repression requires the formation of a multiprotein repression complex at the ventral silencer. *Mol Cell Biol* **18**, 6584-6594.

Von Ohlen, T. and Doe, C. Q. (2000). Convergence of Dorsal, Dpp, and Egfr signaling pathways subdivides the *Drosophila* neuroectoderm into three dorsal-ventral columns. *Dev Biol* **224**, 362-372.

Wang, B., Krall, E. B., Aguirre, A. J., Kim, M., Widlund, H. R., Doshi, M. B., Sicinska, E., Sulahian, R., Goodale, A., Cowley, G. S., et al. (2017). ATXN1L, CIC, and ETS transcription factors modulate sensitivity to MAPK pathway inhibition. *Cell Rep* **18**, 1543-1557.

Wang, X., Goldstein, D., Crowe, P. J. and Yang, J.-L. (2016). Next-generation EGFR/HER tyrosine kinase inhibitors for the treatment of patients with non-small-cell lung cancer harboring EGFR mutations: a review of the evidence. *Onco Targets Ther* **9**, 5461-5473.

Waring, G. (2000). Morphogenesis of the eggshell in *Drosophila*. *Int Rev Cytol* **198**, 67-108.

Weiss, J. B., Von Ohlen, T., Mellerick, D. M., Dressler, G., Doe, C. Q. and Scott, M. P. (1998). Dorsoventral patterning in the *Drosophila* central nervous system: the *intermediate neuroblasts defective* homeobox gene specifies intermediate column identity. *Genes Dev* **12**, 3591-3602.

Whitmarsh, A. J. (2007). Regulation of gene transcription by mitogen-activated protein kinase signaling pathways. *Biochim Biophys Acta* **1773**, 1285-1298.

Wiegmann, B. M., Trautwein, M. D., Winkler, I. S., Barr, N. B., Kim, J.-W., Lambkin, C., Bertone, M. A., Cassel, B. K., Bayless, K. M., Heimberg, A. M., et al. (2011). Episodic radiations in the fly tree of life. *Proc Natl Acad Sci USA* **108**, 5690-5695.

Wilson, M. J. and Dearden, P. K. (2009). Tailless patterning functions are conserved in the honeybee even in the absence of Torso signaling. *Dev Biol* **335**, 276-287.

Yang, L., Paul, S., Trieu, K. G., Dent, L. G., Froldi, F., Forés, M., Webster, K., Siegfried, K. R., Kondo, S., Harvey, K., et al. (2016). Minibrain and Wings apart control organ growth and tissue patterning through down-regulation of Capicua. *Proc Natl Acad Sci USA* **113**, 10583-10588.

Yip, S., Butterfield, Y. S., Morozova, O., Chittaranjan, S., Blough, M. D., An, J., Birol, I., Chesnelong, C., Chiu, R., Chuah, E., et al. (2012). Concurrent *CIC* mutations, *IDH* mutations, and 1p/19q loss distinguish oligodendrogliomas from other cancers. *J Pathol* **226**, 7-16.

Yoshimoto, M., Graham, C., Chilton-MacNeill, S., Lee, E., Shago, M., Squire, J., Zielenska, M. and Somers, G. R. (2009). Detailed cytogenetic and array analysis of pediatric primitive sarcomas reveals a recurrent *CIC-DUX4* fusion gene event. *Cancer Genet Cytogenet* **195**, 1-11.

Zhao, D., Woolner, S. and Bownes, M. (2000). The Mirror transcription factor links signalling pathways in *Drosophila* oogenesis. *Dev Genes Evol* **210**, 449-457.

Zoghbi, H. Y. and Orr, H. T. (2009). Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1. *J Biol Chem* **284**, 7425-7429.

ANEXO

PLOS | GENETICS

RESEARCH ARTICLE

EGFR/Ras Signaling Controls *Drosophila* Intestinal Stem Cell Proliferation via Capicua-Regulated Genes

Yinhua Jin[1], Nati Ha[2], Marta Forés[3], Jinyi Xiang[1], Christine Gläßer[1], Julieta Maldera[1], Gerardo Jiménez[3,4], Bruce A. Edgar[1]*

1 Deutsches Krebsforschungszentrum (DKFZ) - Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH) Allianz, Heidelberg, Germany, **2** Biochemie-Zentrum der Universität Heidelberg (BZH), Heidelberg, Germany, **3** Institut de Biologia Molecular de Barcelona-CSIC, Parc Científic de Barcelona, Barcelona, Spain, **4** Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

* b.edgar@Dkfz-Heidelberg.de

Citation: Jin Y, Ha N, Forés M, Xiang J, Gläßer C, Maldera J, et al. (2015) EGFR/Ras Signaling Controls *Drosophila* Intestinal Stem Cell Proliferation via Capicua-Regulated Genes. PLoS Genet 11(12): e1005634. doi:10.1371/journal.pgen.1005634

Editor: Bruce E. Clurman, Fred Hutchinson Cancer Research Center, UNITED STATES

Received: June 4, 2015

Accepted: October 8, 2015

Published: December 18, 2015

Data Availability Statement: All raw and processed next generation sequencing data are available from the GEO database (accession number GSE74188).

Funding: This work was supported by the DKFZ, DFG grant SFB 873, and ERC Advanced Grant 268515 (http://erc.europa.eu/starting-grants) to BAE. NH was supported by SFB638, SFB1036. MF and GJ were supported by research grants from the Spanish Government (BFU2011-23611) and Fundació La Marató de TV3 (20131730); GJ is an ICREA investigator.The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Epithelial renewal in the *Drosophila* intestine is orchestrated by Intestinal Stem Cells (ISCs). Following damage or stress the intestinal epithelium produces ligands that activate the epidermal growth factor receptor (EGFR) in ISCs. This promotes their growth and division and, thereby, epithelial regeneration. Here we demonstrate that the HMG-box transcriptional repressor, Capicua (Cic), mediates these functions of EGFR signaling. Depleting Cic in ISCs activated them for division, whereas overexpressed Cic inhibited ISC proliferation and midgut regeneration. Epistasis tests showed that Cic acted as an essential downstream effector of EGFR/Ras signaling, and immunofluorescence showed that Cic's nuclear localization was regulated by EGFR signaling. ISC-specific mRNA expression profiling and DNA binding mapping using DamID indicated that Cic represses cell proliferation via direct targets including *string (Cdc25)*, *Cyclin E*, and the ETS domain transcription factors *Ets21C* and *Pointed (pnt)*. *pnt* was required for ISC over-proliferation following Cic depletion, and ectopic *pnt* restored ISC proliferation even in the presence of overexpressed dominant-active Cic. These studies identify Cic, Pnt, and Ets21C as critical downstream effectors of EGFR signaling in *Drosophila* ISCs.

Author Summary

Studies suggest that epidermal growth factor receptor (EGFR) signaling activation is a causal driver of many stem cell-derived epithelial cancers, including colorectal cancer. As in the human intestine, epithelial renewal in *Drosophila* intestine is orchestrated by intestinal stem cells (ISCs). EGFR signaling also plays an important role in regulating ISC proliferation in flies. However, the mechanism by which EGFR/Ras/MAPK signaling promotes ISC proliferation is poorly understood. Here we demonstrate that the transcriptional repressor, Capicua (Cic), mediates these functions of EGFR signaling. We found that the critical role of Cic as a negative regulator of cell proliferation in the fly midgut is consistent

Competing Interests: The authors have declared that no competing interests exist.

with its tumor suppressor function in mammalian cancer development. The direct target genes of Cic were identified by ISC-specific mRNA expression profiling and DNA binding mapping (DamID) method. Cic represses cell proliferation via regulating *string (stg)*, *Cyclin E (CycE)*, and the ETS domain transcription factors *Ets21C* and *pointed* (*pnt*). Using genetic tests we show that these interactions are meaningful for regulating stem cell proliferation. Combining our knowledge of Cic with what was previously known about CIC in tumor development, we propose that human CIC may regulate Ets transcription factors and cell cycle genes in Ras/MAKP-activated tumors.

Introduction

EGFR/Ras/MAPK signaling has diverse functions in regulating cell proliferation, growth, differentiation and survival in most animal cells [1]. Abundant studies also indicate that epidermal growth factor receptor (EGFR) activation is a causal driver of many cancers, including breast, lung, brain, and colorectal cancer [2]. Similarly, activating mutations in KRAS and BRAF, which are essential downstream effectors of the EGFR, are among the most common mutations found in a very wide range of human cancers [3,4]. However, despite much study, many questions remain to be answered to fully understand the impact of EGFR and its downstream effectors during normal cell function and in carcinogenesis. As many epithelial cancers arise through dysregulation of the stem cell self-renewal and homeostatic maintenance of the epithelium [5], understanding the precise functions of EGFR signaling in epithelial homeostasis is very important.

The *Drosophila* midgut is an outstanding model system to study the basis of epithelial homeostasis due to its simple structure, similarity to the mammalian intestine, and powerful genetics. As in the mammalian intestine, epithelial turnover in the fly midgut is carried out through a dynamic process mediated by intestinal stem cells (ISCs). ISCs undergo cell division to renew themselves and give rise to transient cells called enteroblasts (EBs), which can further differentiate into either absorptive enterocytes (ECs) or secretory enteroendocrine (EE) cells. When damaged or aged cells are lost from the fly's gut epithelium, ISCs respond by dividing to replenish the epithelium [6,7,8]. During this response multiple *Drosophila* EGFR ligands, namely *spitz (spi)*, *vein (vn)*, and *keren (krn)* are induced in progenitor cells (EBs and ISCs), visceral muscle (VM) and ECs respectively. Thereby, the EGFR signaling pathway is activated in ISCs. This promotes ISC growth, division and midgut epithelial regeneration [9,10,11]. ISCs defective in EGFR signaling cannot grow or divide, are poorly maintained, and are unable to support midgut epithelial replenishment after enteric infection by the bacteria *Pseudomonas entomophila (P.e.)* [11] or *Erwinia carotovora carotovora 15 (ECC15)* [12]. Interestingly, the critical role of EGFR signaling in the *Drosophila* intestine is consistent with its role during mammalian gut homeostasis and colorectal cancer development [10,11,12,13]. EGFR signaling is required for murine ISC growth [14,15], and the deletion of Lrig1, a negative feedback regulator of EGFR signaling, causes excessive ISC proliferation [16]. Furthermore, adenoma formation in $Apc^{min/+}$ mice was severely impaired in a genetic background with partial loss of function of EGFR ($Egfr^{wa2}$) [17].

Despites its importance, the mechanism by which EGFR/Ras/MAPK signaling promotes ISC proliferation is poorly understood in this cell type. Indeed, despite decades of intensive study, the precise linkage between EGFR/Ras/MAPK signaling and cell growth and division is surprisingly obscure for animal cells in general [3]. Textbook models highlight a prevailing model in which EGFR/Ras signaling controls cell proliferation via a Ras-Myc-CyclinD-Rb

pathway [18,19]. While this may have relevance in some human cancers it is clearly not the case in normal *Drosophila* cells, and so other mechanisms should be sought and characterized.

One potentially important downstream effector of EGFR signaling is the HMG-box transcriptional repressor Capicua (Cic). This highly conserved DNA binding factor has been shown to act downstream of receptor tyrosine kinase (RTK)/Ras/MAPK signaling in *Drosophila* eye and wing imaginal discs, embryos, and ovaries [20,21,22,23] where it regulates diverse RTK-dependent processes including cell proliferation, specification, and pattern formation. Cic orthologs from invertebrate and vertebrate species share two well-conserved regions: the HMG-box, presumed to mediate DNA binding at target promoters [21] and a C-terminal domain [24]. The C-terminal region of *Drosophila* Cic contains a "C1" motif important for repressor activity, and a "C2" motif that functions as a MAPK docking site responsible for downregulation of Cic following the activation of RTK signaling [25]. It has been proposed that MAPK phosphorylates Cic in its C2 motif, and that phosphorylated Cic is either degraded or re-localized to the cytoplasm [25]. Cic downregulation controlled by Torso and EGFR signaling varies in different *Drosophila* tissues [24]. For example, Torso RTK signaling, which also works via the Ras/Raf/MAPK pathway, apparently increases the rate of Capicua degradation by promoting its accumulation in the cytoplasm [26]. EGFR signaling has been reported to regulate Cic protein in distinct ways in different tissues. Wing and eye discs cell clones mutant for *Egfr* or *Ras* showed elevated levels of Cic protein [20,27]. In the ovary, in contrast, Cic protein localized to the cytoplasm in cells in which EGFR signaling was active, but in nuclei in cells in which EGFR signaling was inactive [25]. A recent study suggested that Cic actually undergoes a two-step process in releasing its target gene repression: slower changes in nuclear localization occur after a faster reduction of Cic repressor activity [28]. In cultured human cells, EGF stimulated dissociation of human CIC from importin-α4 (also known as KPNA3), an adaptor required for the nuclear import of many proteins. But full length GFP-CIC was nuclear even after EGF stimulation, and the N-terminal half of the CIC protein was found to be nuclear, even though it does not bind to importin-α4. Hence the biological significance of the CIC:importin association remains unclear [29].

CIC, the human homolog of *Drosophila Cic*, has been implicated in several human diseases including spinocerebellar ataxia type 1 (SCA1) neuropathology, oligodendroglioma (OD) [30] and Ewing-like sarcoma [31]. Human *CIC* is frequently mutated in samples from cancer genome studies such as The Cancer Genomic Atlas (TCGA) (S1 Fig) [32,33]. For instance CIC mutation was reported in 6 out of 7 brain tumors [30], 3 out of 11 breast cancers [34] and 6 out of 72 colorectal cancers [35]. The *Drosophila* work suggests that in these cases CIC loss might have the same downstream consequences (*e.g.* cell transformation) as oncogenic activation of the EGFR, RAS or BRAF, but this has not been rigorously evaluated.

During RNAi screening we discovered that depletion of Cic in *Drosophila's* intestinal stem cells (ISCs) activates these cells for rampant proliferation [11]. Based on previous studies in other fly organs we hypothesized that Cic might act as an obligate repressor downstream of EGFR signaling, itself a central driver of normal ISC proliferation in both flies and mice, as well as in many human colorectal cancers, which are frequently mutant for RAS, BRAF, or CIC. However, until now this hypothesis had not been tested and the underlying mechanisms via which Cic might control ISC proliferation remained undefined. In this report we demonstrate that Cic acts as a critical negative downstream regulator of EGFR signaling to control ISC proliferation. We show that EGFR/Ras activity controls Cic nuclear localization, and we present RNA-Seq and DamID-Seq datasets that together constitute a genome-wide survey of potential Cic target genes in *Drosophila* ISCs. Our analysis indicated that Cic not only directly regulates cell cycle regulators such as *string (cdc25)* and *Cyclin E*, but also the ETS transcription factors *pnt* and *Ets21C*, all of which must be de-repressed to activate ISCs for growth and division.

Results

Cic inactivation promotes ISC proliferation

To investigate a potential role for Cic in regulating ISC proliferation, we used the *esg-Gal4-UAS-2XEYFP; Su(H)GBE-Gal80, tub-Gal80^{ts}* system (henceforth referred as *esg^{ts}; Su(H)-Gal80*) to express *UAS-cic-RNAi* specifically in ISCs. After 4 days of *cic-RNAi* induction, a dramatic increase in the number of YFP positive cells (Fig 1A and 1B) and a large increase in ISC mitoses were observed (Fig 1C). Most of the PH3+ cells were YFP+ [YFP+, PH3+ cells = 99.37% ($n^{midguts}$ = 10 midguts, n^{cells} = 994), YFP-, PH3+ cells = 0.63% (n = 10, n^{cells} = 7)], indicating that Cic regulates ISC proliferation cell autonomously. When we used another ISC-specific driver *Dl^{ts} (tub-Gal80^{ts} UAS-GFP; Dl-Gal4)* to knock down *cic* in ISCs specifically, we not only detected the same overproliferation phenotype (S3A, S3B and S3E Fig) but also found that most of mitotic cells were GFP+ (S3F Fig).

Increased GFP+ cells and mitoses were also noticed when the *esgGal4 UAS-GFP tub-Gal80^{ts}* system (henceforth referred as *esg^{ts}*) was used to express *UAS-cic-RNAi* in ISCs and their undifferentiated daughters, the EBs (S2A–S2B and S3 Figs). To further validate the specificity of this RNAi experiment, GFP-marked ISC clones homozygous for the loss-of-function allele *cic^{fetU6}* [22] were generated using the MARCM system [36] (S2C–S2H Fig). The size of marked ISC clones was quantified at intervals after clone induction by measuring GFP-labeled clone areas. *cic* mutant clones were larger than control clones at all time points assayed (Fig 1D). In addition, the numbers of cells per clone were increased in the *cic* mutant clones (Fig 1E). To further confirm Cic's function in the midgut, we generated viable transheterozygotes using three different loss-of-function alleles of *cic*. *cic^{fetE11}* is a P-element insertion mutant, while both *cic^{fetT6}* and *cic^{fetU6}* are homozygous lethal EMS alleles [22]. In addition to the EGFR-related extra wing vein phenotype reported previously [27], these transheterozygote mutants showed increased mitoses in their midguts (Fig 1I). As the ISCs are the predominant dividing cell type in *Drosophila* midguts, these data further indicate a role for Cic as an obligate repressor of ISC proliferation.

To investigate the respective requirements of Cic in the ISC and EB cell types, the EB-specific driver *Su(H)^{ts} [Su(H)-Gal4,UAS-CD8-GFP; tub-Gal80^{ts}]* was used to knock down *cic* in EBs. Increased mitoses were observed after depleting *cic* in EBs (S3C–S3E Fig). However, in this case only a few GFP^{+} EBs were observed in mitosis, while most of the dividing cells marked by PH3 were GFP-negative (S3F Fig). These GFP-negative mitotic cells are likely ISCs. These data indicated that Cic has both cell autonomous and non-cell autonomous functions in regulating ISC proliferation. In this study we followed up on Cic's cell autonomous effects on ISC proliferation, and the non-cell autonomous effect was not investigated further.

Increased Cic activity inhibits ISC proliferation and midgut epithelial regeneration

To determine whether increased Cic function yields a phenotype similar to that of EGFR loss-of-function, we generated transgenic flies harboring *UAS-cic$^{\Delta C2}$-HA or UAS-cic-HA*. Cic$^{\Delta C2}$ is a Cic derivative carrying a deletion of the MAPK docking site-C2 motif, and has been shown to be a dominant repressor that escapes inactivation by MAPK [25]. Either *cic* or *cic$^{\Delta C2}$* were over-expressed in progenitor cells using *esg^{ts}*, and then the flies were fed *Pseudomonas entomophila (P.e.)* for 12 hours to generate an enteric infection. ISCs from control midguts, which expressed GFP only, showed regeneration-associated proliferation [8]. In contrast both *cic* and *cic$^{\Delta C2}$* overexpressing midguts displayed an inhibition of regeneration after 12 hours *P.e.* infection (Fig 1J). To test if *cic* or *cic$^{\Delta C2}$* overexpression could influence turnover of the midgut

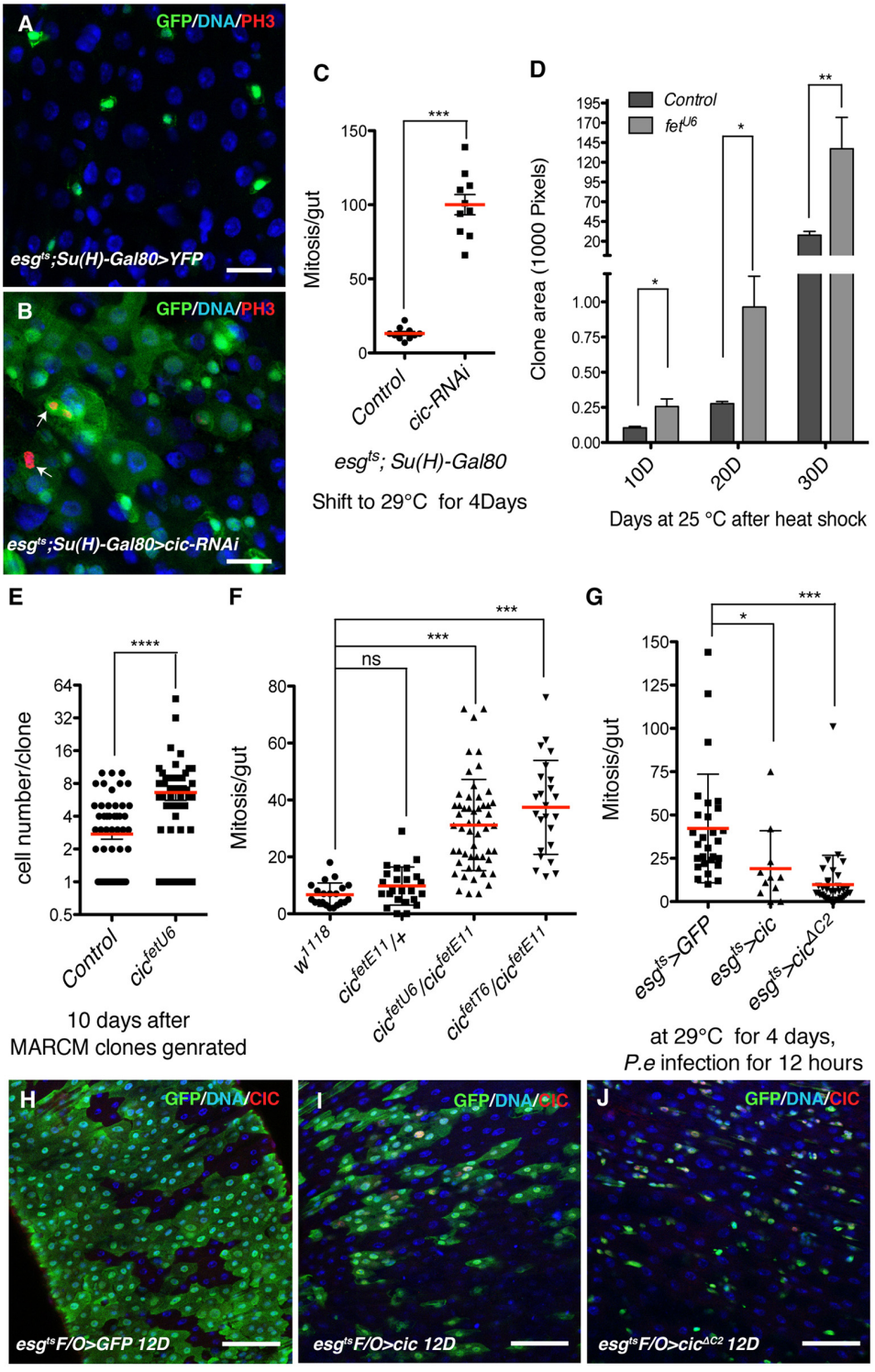


Fig 1. *cic* inactivation promotes ISC proliferation and hyper-activation inhibits ISC proliferation. (A,B) Knock down of Cic in ISCs using the *esg^ts; Su(H)-Gal80* system. ISCs were marked by YFP (green). Samples were stained with anti-PH3 (red) for mitosis and DAPI (blue) for DNA. (A) Control adult midgut (B) Cic knock down midgut after 4 days induction 29°C. Increases in the number of YFP+ cells are observed in *cic* depleted midguts as was a large increase in mitotic cells. (C) Midguts were scored for PH3+ cells after 4 days of

induction of *cic-RNAi*. A strong increase in numbers of ISC mitosis was observed in *cic* knockdown midguts. (D) Clone areas of cic mutant and control WT clones 10, 20, and 30 days after clone induction. Mutant ISCs divided faster and generated bigger clones. (E) Increased number of cells per clone was detected in *cic* mutant clones. Data was quantified 10 days after *cic* mutant clones were generated with the MARCM system. (F) Quantification of pH3-positive cells per adult midgut of the indicated genotype. *cic* transheterozygotes contained significantly more mitotic cells than controls. (G) Quantification of ISC proliferation after 12 hours *P. e*. infection. A decreased number of PH3+ cells, representing dividing ISCs, was observed in midguts overexpressing either *cic* or $cic^{\Delta C2}$ after *P.e*. infection. (H-J) Clones generated by the esg^{ts} *F/O* system are marked with GFP (green), Cic over-expression was confirmed by anti-Cic (red) staining, and nuclei were visualized by DAPI (blue) staining. (E) Control adult midgut 12 days after clone induction (F) midgut overexpressing Cic (G) midgut overexpressing $Cic^{\Delta C2}$ 12 days after clone induction. The size of clones marked by GFP was reduced after Cic or $Cic^{\Delta C2}$ overexpression. Statistical significance was determined by Student's t test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). Error bars represent standard deviations. Scale bars represent 20 μm in A-B and 50 μm in E-G.

doi:10.1371/journal.pgen.1005634.g001

epithelium we used the esg^{ts} *F/O* system *(esg-Gal4; tubGal80^{ts} Act>Cd2>Gal4 UAS-flp UAS-GFP)* [11] to mark all the ISC progeny produced during 12 days of *cic* overexpression. Normally, the posterior midgut epithelium renews it self within about 12 days [8]. Therefore, control midgut epithelia were almost completely replaced by large GFP+ clones that formed during 12 days. However, in the gain-of-function Cic conditions, growth of GFP-marked clones was significantly decreased, indicating that gut epithelial renewal was greatly suppressed (Fig 1H–1J).

Cic regulates ISC proliferation as a downstream effector of EGFR/Ras signaling

EGFR activates ISCs for growth and division via Ras/Raf/MAPK signaling. When an activated form of the EGFR (λTOP) [37] or activated Ras (Ras^{V12G}) [38] is ectopically expressed in progenitor cells, ISC division is dramatically induced. Conversely, EGFR suppression by inducing *Egfr-RNAi*, *Ras-RNAi*, *or MEK-RNAi* in progenitor cells almost completely inhibits ISC division and growth [11,12]. Furthermore, inhibition of EGFR signaling suppresses the activation of ISC divisions after *P.e*. infection [10,11]. As demonstrated above, Cic knockdown and overexpression phenocopy these EGFR overexpression or knockdown phenotypes, respectively, suggesting that Cic may act as a downstream effector in the EGFR signaling in ISCs.

To test the function of Cic in EGFR signaling we performed epistasis tests. After 2 days of clone induction with the esg^{ts} *F/O* system, control midguts generated only 2-cell clones, whereas clones overexpressing an activated variant of the EGFR, (*λtop*), grew very large and showed increased ISC division. However, when *cic* or $cic^{\Delta C2}$ was co-overexpressed along with *λtop*, clone sizes and ISC mitoses were significantly reduced (Fig 2A–2C and 2I). Overexpression of Cic or $Cic^{\Delta C2}$ could also partially inhibit the ISC growth effects of Ras^{V12S35}, an activated allele that can activate RAF/MAPK signaling but not PI3K signaling [38] (Fig 2D, 2H, and 2J). Furthermore, we used esg^{ts} to induce *Egfr-RNAi*, *or Ras-RNAi* in combination with *cic-RNAi*. The *cic*, *Egfr* or *cic*, *Ras* double RNAi animals exhibited increased ISC mitosis relative to controls expressing *Ras-RNAi* or *Egfr-RNAi* only (Fig 2E–2G and 2K), indicating that reduced ISC proliferation caused by the inactivation of EGFR signaling can be restored by Cic knockdown. These epistasis data further support the hypothesis that Cic acts as a negative downstream effector of EGFR to regulate ISC proliferation.

EGFR signaling controls Cic subcellular localization

To understand how EGFR signaling controls Cic in ISCs, we expressed HA-tagged Cic or $Cic^{\Delta C2}$ protein in midgut progenitor cells (ISCs and EBs). As expected, HA-tagged Cic or

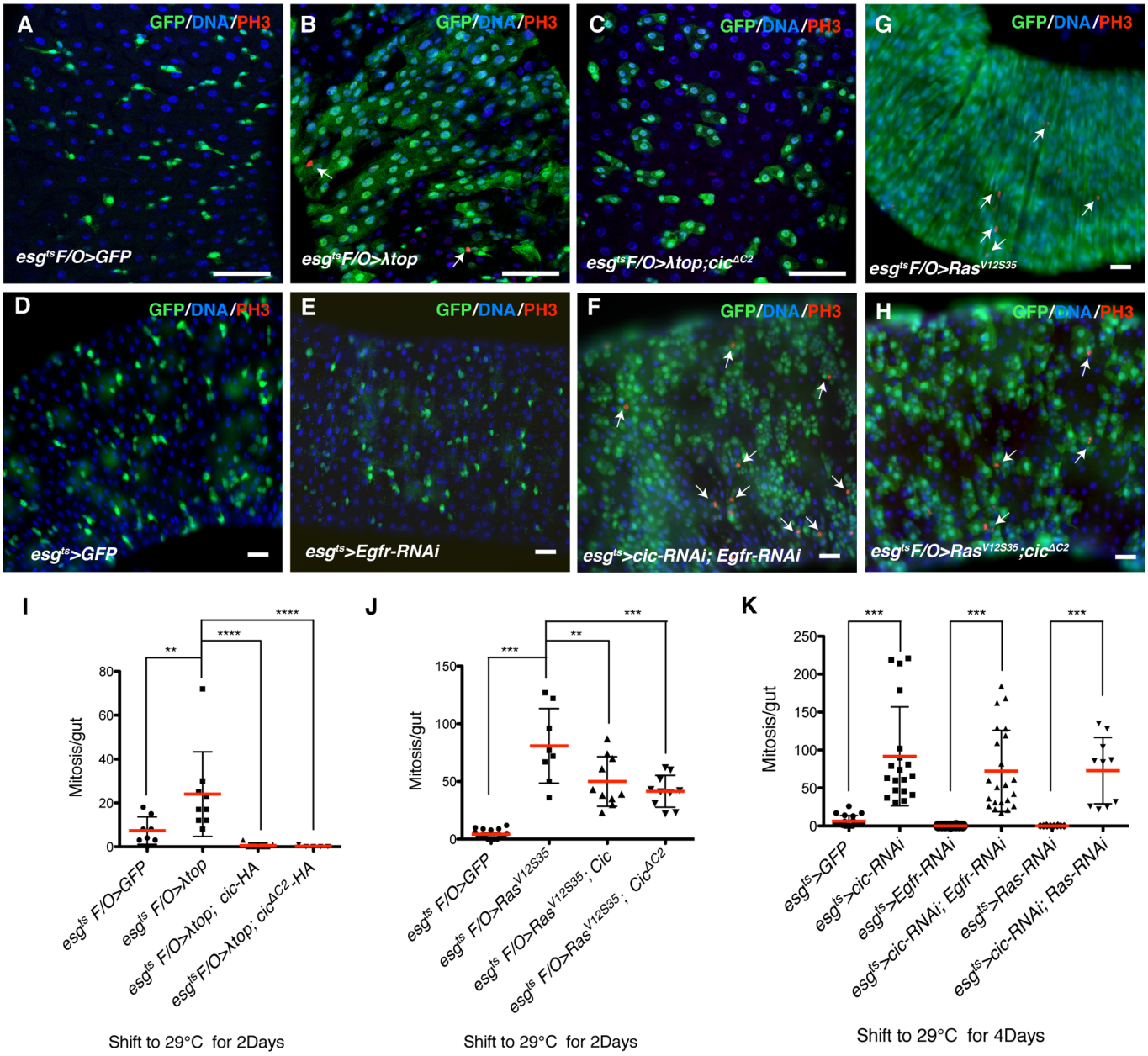


Fig 2. Cic regulates ISC proliferation as a downstream effector of EGFR signaling. (A–C) Results of the *λtop* and *cic* epistasis tests, carried out using the *esg^{ts} F/O* system, to co-express the indicated transgenes with GFP for 2 days at 29°C. (A) Control of adult midgut (B) *λtop* overexpresssing midgut (C) *λtop* and *cic^{ΔC2}* co-overexpresssing midgut. GFP+ clones (green) expressing *λtop* were much smaller when *cic^{ΔC2}* was co-overexpressed. Samples stained with anti-PH3 (red) and DAPI (blue). (D-F) Results of the epistasis test between *cic* and *egfr*, carried out using the *esg^{ts}* system to express the indicated transgenes for 4 days at 29°C. (D) Control adult midgut, (E) *Egfr-RNAi* expressing midgut, (F) *Egfr-RNAi* and *cic-RNAi* co-expressing midgut. The number of GFP+ cells (green) still promoted by depleting *cic* in EGFR/Ras inactivated background. Samples were stained with anti-PH3 (red) and DAPI to visualize nuclei. (G-H) Results of epistasis tests between *Ras^{V12S35}* and *cic*, carried out using the *esg^{ts} F/O* system. The transgenes were induced for 2 days at 29°C (G) *Ras^{V12S35}* over-expressing midgut (H) *Ras^{V12S35}* and *cic^{ΔC2}* co-over expressed midgut. Size of GFP+ clones (green) in *Ras^{V12S35}* and *cic^{ΔC2}* co-overexpressing midgut was significantly reduced. Samples were stained with anti-PH3 (red) and DAPI to visualize nuclei. (I-K) ISC mitoses as quantified by scoring PH3+ cells. (I) Quantification of ISCs mitoses for the *λtop* and *cic* epistasis test. The increase in mitoses induced by *λtop* was completely suppressed by *cic or cic^{ΔC2}* over expression. (J) Quantification of ISC mitoses from *Ras^{V12S35}*/*cic* epistasis tests. The increase in mitosis induced by *Ras^{V12S35}* was partially suppressed by *cic or cic^{ΔC2}* over expression. (K) Quantification of ISCs mitosis in *cic* and either *Egfr or Ras* double knock down midguts. The increase in ISC mitoses induced by *cic-RNAi* is still observed when either *Egfr* or *Ras RNAi* is also expressed. Error bars represent standard deviations. Statistical significance was determined by Student's t test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). Scale bars represent 50μm (A-H).

doi:10.1371/journal.pgen.1005634.g002

PLOS | GENETICS

$Cic^{\Delta C2}$ proteins were only detected in nuclei under normal conditions (Fig 3A–3A' and 3B–3B'). However, HA-tagged Cic protein accumulated nearly exclusively in the cytoplasm when Ras^{V12S35} was co-expressed with it (Fig 3E–3E'). In contrast, $Cic^{\Delta C2}$ remained in the nucleus even following ectopic Ras^{V12S35} expression (Fig 3F–3F'). A similar but milder re-localization of Cic protein from the nucleus to the cytoplasm was observed following *P.e.* infection (Fig 3C and 3C'), a treatment known to increase MAPK signaling in the gut [11]. It is interesting to note that $Cic^{\Delta C2}$ did not completely suppress Ras^{V12S35} induced ISC proliferation, even though it remained localized to nuclei in Ras^{V12S35} expressing cells (Fig 2H and 2J). However, nuclear $Cic^{\Delta C2}$ lost its characteristic punctate localization in the presence of Ras^{V12S35} expression, and became more diffusely localized in the nucleoplasm (Fig 3F–3F'). These results suggest that, although EGFR signaling controls Cic nucleo-cytoplamic localization via the C2 motif, there may be a second MAPK-dependent mechanism to regulate Cic repressor activity, involving dissociation from chromatin, that is C2-independent.

Cic represses cell cycle genes in ISCs

Cic has been studied in several cell types from both *Drosophila* and humans. In human melanoma cells, CIC represses mRNA expression of the PEA3 subfamily of ETS transcription factors, namely ETV1, ETV4 and ETV5 [29]. In early *Drosophila* development post-transcriptional down-regulation of Cic by the Torso and EGFR pathways regulates terminal and dorsal-ventral patterning, respectively, by allowing expression of Cic target genes such as *huckebein* (*hkb*), intermediate *neuroblasts defective* (*ind*), and *argos* (*aos*) [39]. However, a genome-wide mapping of Cic target genes has not yet been reported.

To identify Cic target genes involved in ISC growth and proliferation we profiled Cic binding throughout the genome using the "TaDa" (Targeted DamID)" technique. The TaDa method involves low-level expression of a GAL4-inducible Dam methylase-fusion protein in a specific cell type, enabling cell-specific profiling without cell isolation [40,41]. Here, we induced a low level of Dam-only or Dam-Cic fusion protein in progenitor cells (ISC & EB) using the esg^{ts} system and a 24-hour induction. Genomic DNA was extracted from isolated midguts, digested with Dpn I, which cuts only methylated GATCs, and amplified. The amplified gDNA fragments were subjected to high-throughput sequencing, rather than tiling microarrays as previously reported [40]. We identified 2279 binding sites that were highly enriched (log2 fold change $>$ 3, false discovery rate$<$0.01%) when comparing Dam-Cic to Dam alone samples (S1 Table). These sites were non-randomly distributed in the genome, and were significantly over-represented ~500 bp 5' to Transcription Start Sites (TSS; Fig 4A). Cic DamID was also performed on progenitor cells from *P.e.* infected midguts. After a 24 hours induction of Dam or Dam-Cic transgenes via the esg^{ts} system, flies were fed *P.e.* bacteria for 16 hours. The number of highly enriched (log2 fold change $>$ 3, FDR $<$ 0.1%) peaks was reduced to 1903. In addition, the fold change of peaks (Dam-Cic vs Dam-alone) after *P.e.* infection was significantly decreased (Figs 4B–4C and S4A). The frequency of peaks adjacent to TSS was also significantly reduced in the *P.e.*-infected midgut sample (Fig 4A). We believe that this decrease was due to the change of Cic localization from the nucleus to cytoplasm, which was caused by the activation of EGFR/Ras/MAPK signaling after infection.

To further understand how Cic regulates ISC proliferation we performed gene expression profiling using amplified mRNA from FACS-sorted esg+ progenitor cells that expressed *cic-RNAi*, and controls. As a way to identify potentially direct target genes of Cic, the RNA-Seq and DamID-Seq data sets were cross-compared. Amongst 439 transcriptionally up-regulated genes (>1.5 fold change, 90% CI) (S2 Table), a large fraction [134 genes, (S3 Table)] had Cic binding sites as defined by DamID (Fig 4E). We next examined the enrichment of the DamID

PLOS | GENETICS

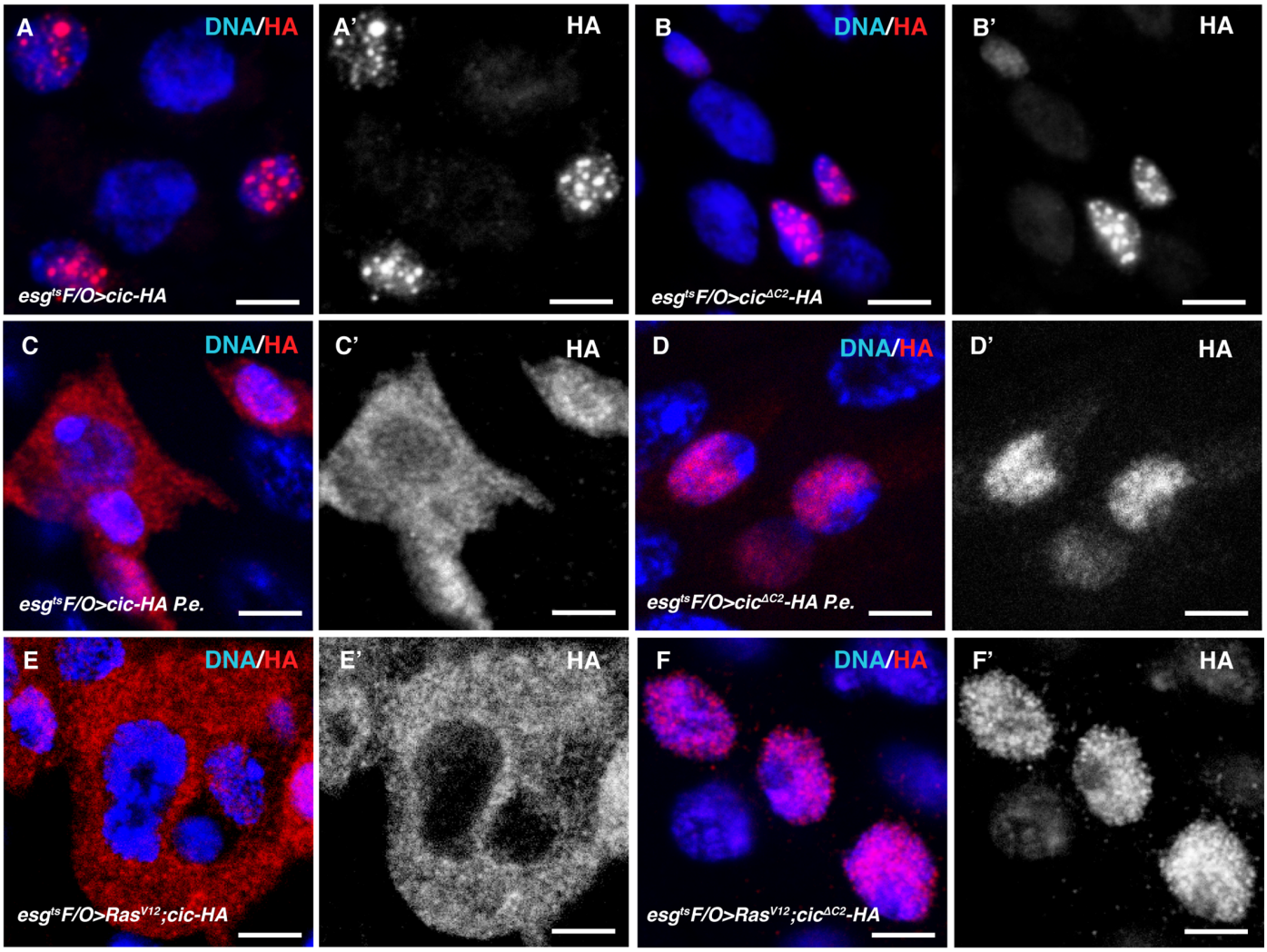
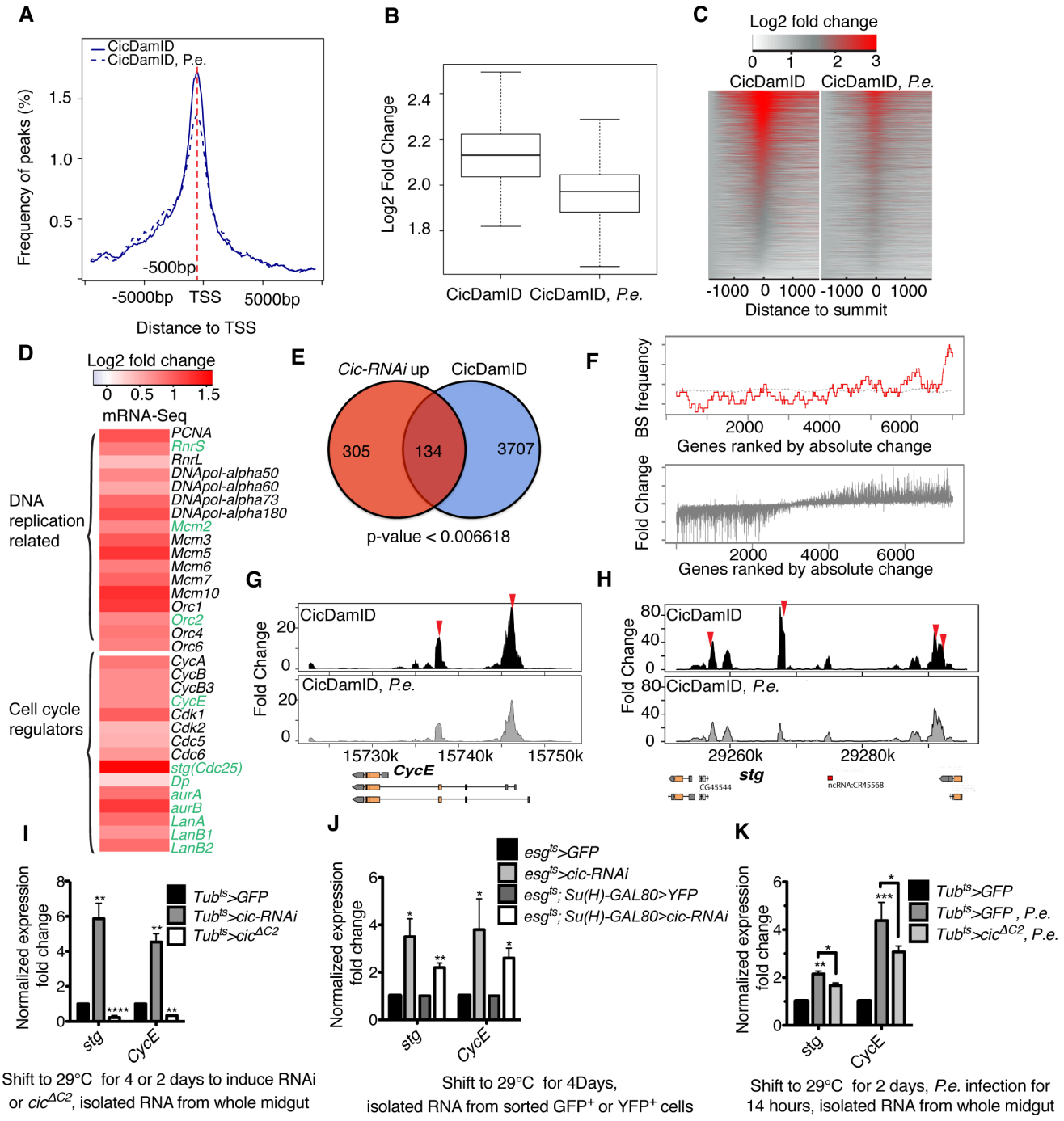


Fig 3. EGFR signaling controls Cic subcellular localization in ISCs. (A, B) Cic and Cic$^{\Delta C2}$ localized in the nuclei. Transgene expression was induced using the *esg^{ts} F/O* system at 29°C for 2 days. HA-tagged Cic or Cic$^{\Delta C2}$ protein was detected by anti-HA antibody (red). Nuclear DNA is marked by DAPI staining (blue) (A) *cic-HA* overexpressing midgut. (B) *cic$^{\Delta C2}$-HA* overexpressing midgut. (C, D) Cic but not Cic$^{\Delta C2}$ protein accumulated in the cytoplasm after *P.e*. infection. (C) *cic-HA* overexpressing midgut, exposed to *P.e*. bacteria for 16 hours. (D) *cic$^{\Delta C2}$-HA* overexpressing midgut after 16 hours *P.e*. infection. (E, F) Cic protein accumulated in the cytoplasm when EGFR signaling was activated by *Ras^{V12S35}*. (E) *Ras^{V12S35}* and *cic-HA* overexpressing midgut. (F) *Ras^{V12S35}* and *cic$^{\Delta C2}$-HA* overexpressing midgut. Cic $^{\Delta C2}$ proteins stayed in the nucleus even after overexpressing *Ras^{V12S35}* to activate MAPK signaling. Scale bars represent 5μm.

doi:10.1371/journal.pgen.1005634.g003

peaks in the transcriptionally induced genes, ranked by absolute expression change in *cic* knockdown progenitor cells (see Materials and Methods). Cic binding peaks that were significantly reduced upon *P.e*. infection (< 2 fold change) were enriched in up-regulated genes from the RNA-Seq dataset (Fig 4F). Hence, the set of genes present in the overlapping set are likely to be direct target genes of Cic. Many cell cycle regulators and genes involved in DNA replication were upregulated in Cic-depleted progenitor cells (Fig 4D). In addition, a large portion of cell cycle control genes that were upregulated upon *cic-RNAi*, including *string (stg, Cdc25)* and *Cyclin E (CycE)*, had Cic binding sites (Fig 4D, 4G, and 4H). To further assess the reliability of this approach we examined the occupancy of Cic on its previously characterized direct target gene-*aos* [39]. Our DamID-Seq data showed that *aos* contained two Cic binding sites within its

PLOS | GENETICS



Shift to 29°C for 4 or 2 days to induce RNAi or $cic^{\Delta C2}$, isolated RNA from whole midgut

Shift to 29°C for 4Days, isolated RNA from sorted GFP^+ or YFP^+ cells

Shift to 29°C for 2 days, *P.e.* infection for 14 hours, isolated RNA from whole midgut

Fig 4. Cic targets genes in ISCs found by DamID-Seq. (A) Graph showing the location of Cic binding relative to annotated transcript TSSs. The distance is from the summit of the Cic peaks to the nearest TSS. Dashed red line showed the summit of the graph is 500bp away from TSS. (B) Box plot showing fold change of peaks in CicDamID and *P.e.* infected CicDam. (C) Heatmap showing fold enrichment of Cic peaks from Cic DamID-Seq without or with *P.e.* infection. Y axis represents genes associated with the Cic binding peaks. (D) Expression heatmap of cell cycle regulators and DNA replication related genes from RNA-Seq data from *cic-RNAi* expressing FACS sorted progenitor cells. The names of the genes that had Cic binding sites by DamID are written in green. (E) Venn diagram showing the overlap between genes upregulated > 1.5 fold upon *cic-RNAi* (left) and genes associated with Cic binding peaks (right) in ISC/EBs. (F) Graph showing correlation between genes upregulated in Cic-depleted progenitor cells, and the Cic-DamID peaks that changed significantly upon *P.e.* infection (upper panel). Lower panel show genes ranked by absolute expression change, and then plotted for expression fold change (bottom). (G, H) Cic binding sites in the *CycE* and *stg* loci, as determined by Cic DamID-Seq in ISC/EBs from control (above) and *P.e.* infected (below) midguts. Vertical bars represent the log2 ratio of the Dam-fusion signal to the Dam-only signal. Red arrows indicate TGAATG(G/A)A motifs. (I-K) mRNA level fold changes of

stg and *CycE* analyzed by qRT-PCR and normalized to *β-Tub* and *Rp49*. (I) *stg* and *CycE* fold enrichment from whole midguts after knocking down or over expressing *cic* in all cells using the *tub*^{ts} *(tubGal4; tubGal80ts)* driver. Transcription of both *stg* and *CycE* was induced in *cic* knock-down midguts and inhibitied in *cic* over-expressing midgut. (J) *stg* and *CycE* expression is upregulated in *cic*-depleted, FACS-sorted progenitor cells (ISC &EB) and ISCs. (K) *stg* and *CycE* expression fold change in *cic* over expressing midguts after *P.e*. infection. The induction of *stg* and *CycE* by *P.e*. infection was suppressed by *cic*$^{\Delta C2}$ overexpression. Statistical significance was determined by Student's t test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). Error bars in each graph represent standard deviations.

doi:10.1371/journal.pgen.1005634.g004

enhancer, and that their occupancy was significantly reduced after *P.e*. infection (S4B Fig). The significant induction of *aos* transcription was verified both by RNA-Seq and qRT-PCR data from FACS-sorted progenitor cells expressing *cic-RNAi* (S4C Fig).

Having confirmed the reliability of our approach for identifying genes that are repressed by Cic in ISCs, we focused on genes likely to contribute to ISC proliferation. We were interested in *stg* and *CycE* because they are transcriptionally induced in proliferating ISCs [42], required for ISC divisions, and also sufficient to induce sustained ISC division when co-overexpresssed [42]. To further test whether Cic regulates the transcription of *stg* and *CycE* we measured their normalized expression ratios in gain- or loss-of-function Cic midguts via RT-qPCR (Fig 4I–4K). The *stg* and *CycE* mRNAs were significantly increased in Cic-depleted midguts, and decreased in midguts expressing the dominant active Cic$^{\Delta C2}$. Strong inductions of *stg* and *CycE* were also observed in Cic-depleted progenitor cells or ISCs purified using FACS (Fig 4J). Moreover, both the *stg* and *CycE* loci had multiple strong Cic-Dam-ID binding peaks containing TGAATG(G/A)A motifs, and binding these peaks were reduced by *P.e*. infection (Fig 4G and 4H). Consistently, the induction of *stg* and *CycE* transcription upon *P.e*. infection was significantly repressed by Cic$^{\Delta C2}$ overexpression (Fig 4K). These data support the notion that Cic controls ISC cell cycle progression by directly repressing transcription of *stg* and *CycE* via binding sites in their regulatory regions.

Cic represses *pnt* and *Ets21C*

It has been suggested that Cic might regulate the transcription of certain members in a subfamily of ETS transcription factors [29,31]. Consistent with this, we identified the *Drosophila* ETS transcription factors *pnt* and *Ets21C* as potential *Cic* direct target genes by both RNA-Seq and DamID-Seq (Figs 5 and S5). These genes contain Cic binding sites, were highly expressed in midgut progenitor cells, and were significantly induced upon infection or *cic* depletion or mutation. Notably, induction of *pnt* and *Ets21C* was detected in FACS-sorted ISCs depleted of Cic (Fig 5B). Moreover, the induction of *pnt* and *Ets21C* expression by *P.e*. infection was suppressed when the dominant active form, Cic$^{\Delta C2}$ was overexpressed (Fig 5D). Similar effects were observed when Cic was either depleted or overexpressed in whole midgut samples (Figs 5C and S5C). These data suggest that Cic also regulates *pnt* and *Ets21C* transcription in *Drosophila* midgut ISCs, by directly binding to these loci. As in the case of *stg* and *CycE*, this regulation appeared to be modulated by *P.e*. infection, most likely in a MAPK-dependent manner.

Cic represses *pnt* via a TGAATGAA motif

The HMG box of Human Cic binds to TGAATG(G/A)A octamers *in vitro* [31]. This motif was also verified as a Cic binding sequence in several Cic target genes in *Drosophila* embryos and wing discs [39]. Notably, the TGAATG(G/A)A motif was observed in 692/2279 Cic binding sites in our DamID-Seq dataset (p-value = 3.045475×10^{-11}). Each of the four Cic target genes discussed above contained more than one TGAATG(G/A)A motifs in its Cic binding sites. Moreover, TGAATGAA motifs found in the *pnt* locus also mapped to Cic binding sites that we determined from *Drosophila* embryo ChIP-Seq (Fig 5E). This suggests that Cic may bind to the

PLOS | GENETICS

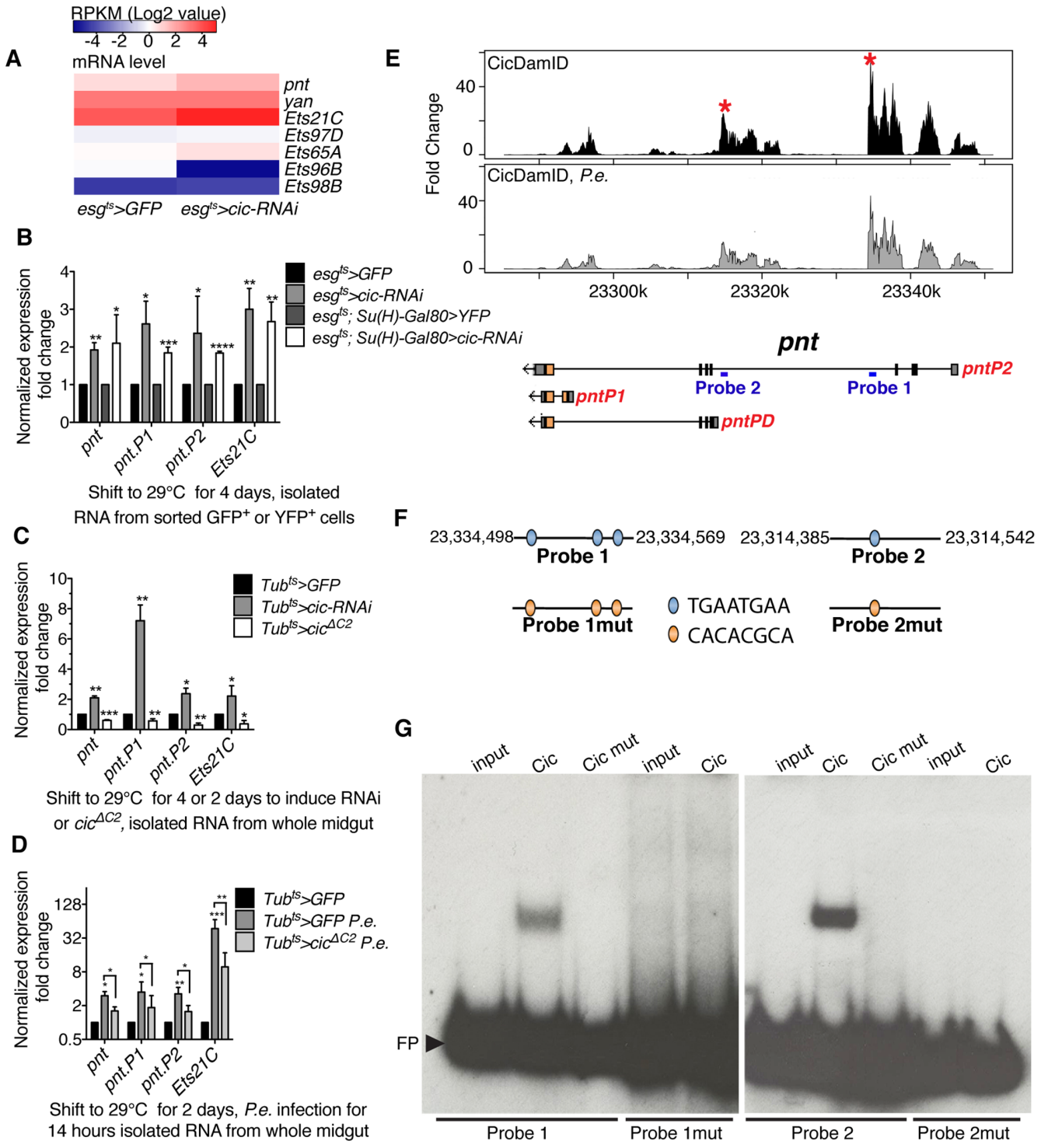


Fig 5. Cic regulates Pnt expression through binding to its genomic locus. (A) Heatmap of mRNA levels indicating RPKM values from RNA-Seq data from control and *cic-RNAi* expressing, FACS-sorted progenitor cells. (B-D) Relative expression of *pnt*, *pntP1*, *pntP2* and *Ets21c* as analyzed by qRT-PCR and normalized to *β-Tub* and *Rp49*. (B) Fold change of expression from the *cic* depleted FACS-sorted progenitor cells and ISCs. (C) *pnt*, *pntP1*, *pntP2* and Ets21C were upregulated in *cic* knock down midguts and downregulated in *cic* overexpressing midguts. (D) Expression change in *cic* overexpressed midgut after *P.e.* infection. Error bars represent standard deviation. Statistical significance was determined by Student's t test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). (E) Cic binding sites in the *pnt* locus, as determined by Cic DamID-Seq of esg+ progenitor cells from control and *P.e.* infected midguts.

Vertical bars represent the log2 ratio of the Dam-fusion signal to the Dam-only signal. Peaks also found in Cic ChIP-Seq from embryos are marked by asterisks. Positions of EMSA probes from the *pnt* locus are indicated by blue bars. (F) Diagram of probes containing TGAATGAA sites. These sites were replaced with other sequences in probes 1 and 2 to generate probes 1mut and 2mut. (G) DNA binding of Cic and HMG-box mutated Cic to probe 1 or 1mut (left panel). DNA binding of Cic and HMG-box mutated Cic to probe 2 or probe 2mut (right panel). FP indicates "free probes."

doi:10.1371/journal.pgen.1005634.g005

pnt locus via TGAATGAA octamers, and that the occupancy of Cic at the *pnt* locus may also be conserved in different *Drosophila* cell types. To further evaluate this hypothesis we performed electrophoretic mobility shift assay (EMSA). Cic showed specific binding to two DNA fragments from the *pnt* locus that were identified as prominent *in vivo* Cic binding peaks by DamID-Seq and ChiP-Seq (Fig 5F and 5G). Importantly, the EMSA interaction was lost when the HMG-box in Cic was mutated, or when the TGAATGAA motifs were mutated. These data strongly support the idea that Cic directly regulates *pnt* transcription by directly binding to TGAATGAA motif in *pnt* locus.

Pnt regulates ISC proliferation as a direct target of Cic

Pnt is believed to be a downstream effector of EGFR signaling in developing *Drosophila* eyes [43,44,45]. The *pnt* locus produces two alternative transcripts that encode two different protein isoforms: PNTP1 and PNTP2 [44]. PNTP1 was proposed to be a constitutive activator of transcription, whereas PNTP2 has a PNT (pointed) domain that was reported to be phosphorylated by MAP kinase *in vitro* [45]. The mutant protein, $PNTP2^{T151A}$, which cannot be phosphorylated *in vitro*, was unable to rescue *pnt* phenotype in eyes but instead enhanced the mutant phenotype, suggesting that the PNT domain is an auto-inhibitory domain that can be inactivated by MAPK-dependent phosphorylation [45]. Furthermore PNTP2 is thought to induce transcription of PNTP1, which might thereby encode the final nuclear effector of the EGFR pathway in eye discs [43]. In the midgut, we found an interesting interaction between Pnt and Cic: *pntP1* and *pntP2* were both induced when Cic was depleted, and both decreased when Cic was overexpressed (Figs 5B, 5C, and S5A). The expression of transcripts encoding both isoforms was also increased in *P.e.* infected guts (Figs 5D and S5B). This raises the possibility that *pnt* might be an important downstream effector of Cic in controlling ISC proliferation. To test this we over-expressed either *pntP1* or *pntP2* in progenitor cells using the esg^{ts} or Dl^{ts} driver systems. After 4 days of transgene induction a dramatic increase in ISC division was evident in response to either *pntP1* or *pntP2* (Figs 6A–6B, 6I, and S6A–S6B). Conversely, mutant clones that were generated using a *pnt* null allele ($pnt^{\Delta 88}$) [46] did not grow past the 2-cell stage (S6D Fig). Moreover, when we depleted *pnt* in progenitor cells by expressing a *pnt-RNAi* that recognizes both isoforms, or generated homozygous *pnt* null mutant ISCs via MARCM, ISC proliferation after *P.e.* infection was suppressed (Figs 6C–6D, 6J, and S6E). Next, we investigated the functional significance of the inhibition of *pnt* expression by Cic. Whereas loss of Cic function induced massive ISC proliferation, inhibiting both isoforms of *pnt* in this context suppressed this over-proliferation (Figs 6G–6H, 6K and S6F–S6G). Conversely, when we over-expressed either *pntP1* or *pntP2* in ISCs that also overexpressed $Cic^{\Delta C2}$, the inhibitory effect of $Cic^{\Delta C2}$ on proliferation was bypassed and the cells divided (Figs 6F, 6L and S6C). Hence, a significant fraction of the ISC over proliferation caused by Cic knockdown can be attributed to Cic's effects on *pntP1 and pntP2*

Interestingly, mutant clones generated using a *pntP1* specific mutant allele, $pnt^{\Delta 33}$ [45,47], or a *pntP2* specific mutant allele, $pnt^{\Delta 78}$ [45,47], grew normally. However ISCs mutant for the *pnt* null allele $pnt^{\Delta 88}$ did not expand (S6D Fig). In addition, $pnt^{\Delta 33}$ and $pnt^{\Delta 78}$ homozygous clones in which *cic* was depleted by RNAi had similar numbers of cells to *cic*-depleted control clones (*i.e.* they overgrew), whereas $pnt^{\Delta 88}$ null mutant clones contained significantly fewer

PLOS | GENETICS

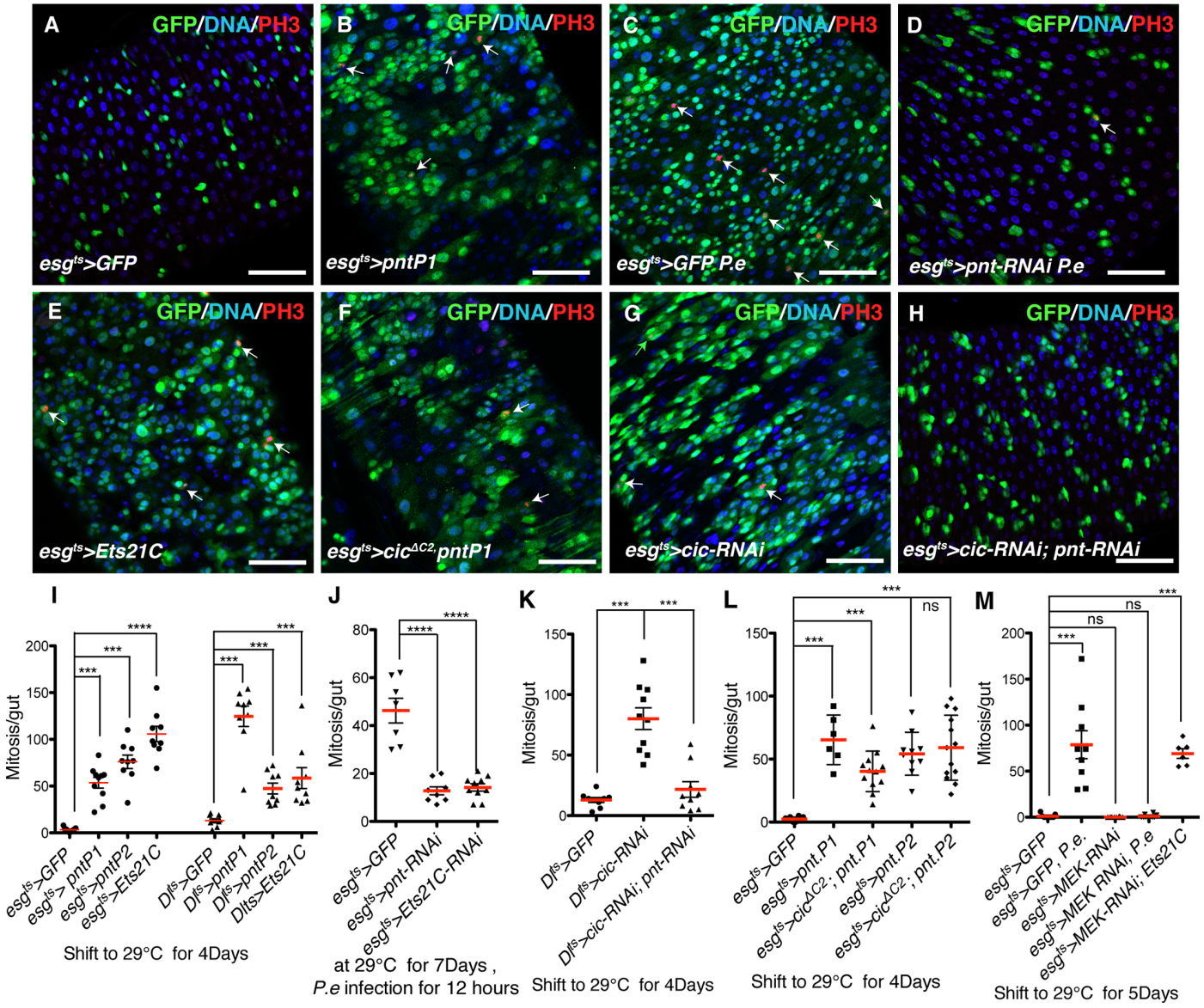


Fig 6. Cic controls ISC proliferation by regulating *pnt* transcription. (A–E) Effects caused by *pntP1* overexpression and RNAi's. Transgenes were induced using the *esg*^{ts} system at 29°C for 4 days, and samples were stained for GFP (green), DNA (blue) and mitoses (PH3, red). White arrows pointing out PH3 signals. (A) Control adult midgut. (B) *pntP1* overexpressing midgut after 4 days induction at 29°C. (C) Control midgut after 12 hours *P.e*. infection. (D) *pnt* knockdown midgut after 12 hours *P.e*. infection. Fewer GFP+ and PH3+ cells are observed. (E) *Ets21C* overexpressing midgut, showing more PH3 + ISCs (arrows) and GFP+ ISCs and EBs (green). (F-H) Ectopic expression or loss of *pnt* bypasses ISC phenotypes caused by *cic* overexpression or depletion. (F) *pnt* and *cic*$^{\Delta C2}$ co-over-expressing midgut after 4 days induction at 29°C. GFP+ progenitor cells were still able to proliferate. (G) *cic* knockdown adult midgut and (H) *pnt*, *cic* double knockdown midgut. The increased number of progenitor cells marked by GFP upon *cic* knockdown was decreased by also knocking down *pnt*. (I-L) Quantification of PH3+ cells in adult midguts of the indicated genotypes. (I) *pntP1*, *pntP2* or *Ets21C* overexpression driven by *esg*^{ts} or *Dl*^{ts}. All the *pntP1*, *pntP2* and *Ets21C* overexpressing midguts contained more dividing ISCs. (J) *pnt* or *Ets21C* knockdown midguts after *P.e*. infection. ISC mitoses caused by *P.e*. infection were reduced in *pnt* or *Ets21C* knockdown midguts. (K) *pnt* and *cic* knock down using *Dl*^{ts} system. Fewer mitotic ISCs were observed in the *pnt* and *cic* double knockdown midgut than the *cic* knockdown midgut. (L) *pnt* and *cic*$^{\Delta C2}$ co-overexpressing midguts. *cic*$^{\Delta C2}$ overexpression could not inhibit ISC mitoses caused by *pnt* overexpression. (M) Quantification of PH3+ cells from adult midguts following *P.e*. infection. *MEK-RNAi* completely blocked infection-driven ISC mitoses, but could not inhibit ISC proliferation driven by overexpressed Ets21c. Statistical significance was determined by Student's t test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). Error bars in each graph represent standard deviation. Scale bars represent 50μm.

doi:10.1371/journal.pgen.1005634.g006

cells (S6G Fig). These data not only support our conclusion that *pnt* is required for ISC proliferation as a target of Cic, but show that PNTP1 and PNTP2 have redundant function in regulating ISC proliferation. Furthermore, *pntP2* homozygous mutant ISCs did not appear to have any defect in proliferation upon *P.e.* infection (S6E Fig). Overall these results indicate that *pntP2*, the isoform proposed to be activated directly by MAKP phosphorylation [45], is not specifically required in ISC proliferation.

Pnt is the *Drosophila* ortholog of the human ETS2 transcription factor and has a conserved ETS-type DNA binding domain, while Ets21C is the *Drosophila* ortholog of the human proto-oncogene ERG. In addition to having Cic binding sites, RT-PCR and RNA-Seq data showed that Ets21C was highly induced upon *P.e.* infection (Figs 5A and S5C). Moreover RNAi mediated depletion experiments indicated that Ets21C was also required for ISC proliferation in response to *P.e.* infection (Fig 6J). Over-expression of *Ets21C* caused a strong increase of ISC division (Fig 6E and 6I) suggesting that transcriptional induction of *Ets21C* could promote ISC proliferation. Furthermore, ectopic expression of *Ets21C* in progenitor cells could bypass the strong growth-suppressive effect of depleting MEK (Fig 6M). These data indicated that Cic controls ISC proliferation in part by regulating *Ets21C* transcription.

Finally, we tested whether Yan, an inhibitory ETS type transcription factor, reported to be MAPK responsive and to compete with Pointed for binding to common sites on the DNA [45,48,49], had an opposite function in ISCs. Although *yan* mRNA is expressed in the midgut (Fig 5A), *yan* depletion from ISCs did not produce a detectable effect (S6F Fig). Two independent *yan-RNAi* lines were used, both of which were proven to be effective by qRT-PCR (S6H Fig). In summary these observations suggest that EGFR signaling controls ISC growth and division by regulating the activity of Cic, Pnt and Ets21C but not Yan, and that Cic directly represses *pntP1*, *pntP2* and *Ets21C* in this context.

Discussion

It is well established that EGFR signaling is essential to drive ISC growth and division in the fly midgut [10,11,12]. However, the precise mechanism via which this signal transduction pathway activates ISCs has remained a matter of inference from experiments with other cell types. Moreover, despite a vast literature on the pathway and ubiquitous coverage in molecular biology textbooks, the mechanisms of action of the pathway downstream of the MAPK are not well understood for any cell type. From this study, we propose a model summarized in Fig 7. Multiple EGFR ligands and Rhomboid proteases are induced in the midgut upon epithelial damage, which results in the activation of the EGFR, RAS, RAF, MEK, and MAPK in ISCs. MAPK phosphorylates Cic in the nucleus, which causes it to dissociate from regulatory sites on its target genes and also translocate to the cytoplasm. This allows the de-repression of target genes, which may then be activated for transcription by factors already present in the ISCs. The critical Cic target genes we identified include the cell cycle regulators *stg (Cdc25)* and *Cyclin E*, which in combination are sufficient to drive dormant ISCs through S and M phases, and *pnt* and *Ets21C*, ETS-type transcriptional activators that are required and sufficient for ISC activation.

Although we found more than 2000 Cic binding sites in the ISC genome, not all of the genes associated with these sites were significantly upregulated, as assayed by RNA-Seq, upon Cic depletion. One possible explanation for this is that Cic binding sites from DamID-Seq were also associated with other types of transcription units (miRNAs, snRNAs, tRNAs, rRNAs, lncRNAs) that were not scored for activation by our RNA-Seq analysis. Indeed a survey of the Cic binding site distributions suggests this (S5 Table). This might classify some binding sites as non-mRNA-associated. However, it is also possible that many Cic target genes may require

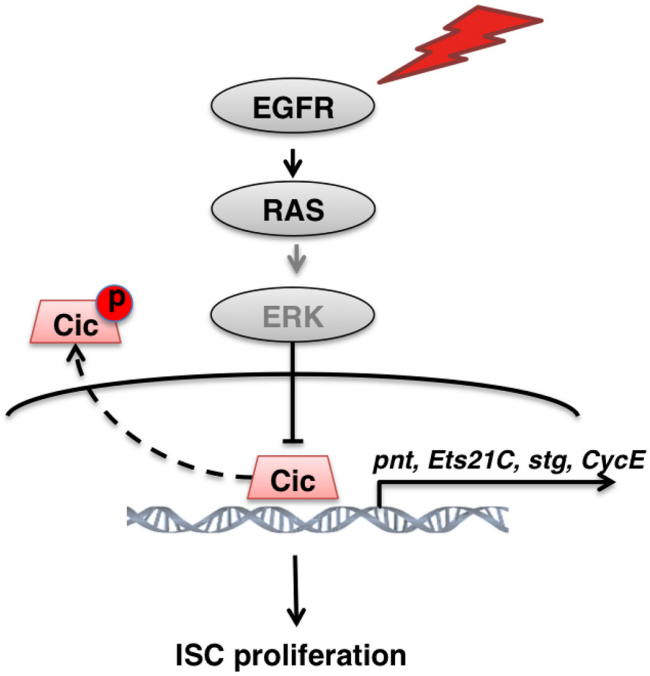


Fig 7. Model for Cic control of Drosophila ISC proliferation. Upon damage, activated EGFR signaling mediates activation of ERK, which phosphorylates Cic, and relocates it to the cytoplasm. As a result, *stg*, *CycE*, *Ets21C* and *pnt* transcription are relieved from Cic repression, and induce ISC proliferation.

doi:10.1371/journal.pgen.1005634.g007

activating transcription factors that are not expressed in ISCs. Such genes might not be strongly de-repressed in the gut upon Cic depletion.

In other *Drosophila* cells MAPK phosphorylation is thought to directly inactivate the ETS domain repressor Yan, and to directly activate the ETS domain transcriptional activator Pointed P2 (PNTP2) [45,50]. In fact Pnt and Yan have been shown to compete for common DNA binding sites on their target genes [45,48,49]. Thus, previous studies proposed a model of transcriptional control by MAPK based solely on post-translational control of the activity of these ETS factors. However, we found that Pnt and Ets21C were transcriptionally induced by MAPK signaling, and could activate ISCs if overexpressed, and that depleting *yan* or *pntP2* had no detectable proliferation phenotype. In addition, overexpression of PNTP2 was sufficient to trigger ISC proliferation, suggesting either that basal MAPK activity is sufficient for its post-translational activation, or that PNTP2 phosphorylation is not obligatory for activity. Moreover, *pntP2* loss of function mutant ISC clones had no deficiency in growth (S6D Fig) even after inducing proliferation by *P.e.* infection, which increases MAPK signaling (S6E Fig). These observations indicate that the direct MAPK→PNTP2 phospho-activation pathway is not uniquely or specifically required for ISC proliferation. Or results suggest instead that transcriptional activation of *pnt* and *Ets21c* via MAPK-dependent loss of Cic-mediated repression is the predominant mode of downstream regulation by MAPK in midgut ISCs.

In addition to activating ISCs for division, EGFR signaling activates them for growth. Previous studies showed loss of EGFR signaling prevented ISC growth and division, and that ectopic Ras^{V12} expression could accelerate the growth not only of ISCs but also post-mitotic enteroblasts [11]. Similarly, our study shows that loss of *cic* caused ISC clones to grow faster than controls, by increasing cell number as well as cell size (Figs 1H and S2C–S2H). For instance, increased size of GFP^+ ISCs and EBs was observed when *cic-RNAi* was induced by the esg^{ts} *or*

esg^{ts}F/O systems (Figs 1B, 6G and S2B). Therefore, in our search for Cic target genes we specifically checked probable growth regulatory genes such as Myc, Cyclin D, the Insulin/TOR components InR, PI3K, S6K and Rheb, Hpo pathway components, and the loci encoding rRNA, tRNAs and snRNAs. We found that Cic bound to the InR, Akt1, Rheb, Src42A and Yki loci. However, of these only InR mRNA was significantly upregulated in Cic-depleted progenitor cells (S4 Table). In surveying the non-protein coding genome, we found that Cic had binding sites in many loci encoding tRNA, snRNA, snoRNA and other non-coding RNAs (S5 Table), though not in the 28S rRNA or 5S rRNA genes (S4 Table). Due to the method we used for RNA-Seq library production, our RNA expression profiling experiments could not detect expression of these loci, and so it remains to be tested whether Cic may regulate some of those non-coding RNA's transcription to control cell growth. It is also possible that Cic controls cell growth regulatory target genes indirectly, for instance via its targets Ets21C and Pnt, which are also strong growth promoters in the midgut (Figs 6A–6B, 6E and S6A–S6B). But given that no conclusive model can be drawn from our data regarding how Cic restrains growth, one must consider the possibility that ERK signaling stimulates cell growth via non-transcriptional mechanisms, as proposed by several studies [51,52,53,54].

The critical role of Cic as a negative regulator of cell proliferation in the fly midgut is consistent with its tumor suppressor function in mammalian cancer development (S1 Fig). Also consistent with our findings are the observations that the ETS transcription factors ETV1 and ETV5 are upregulated in sarcomas that express CIC-DUX, an oncogenic fusion protein that functions as a transcriptional activator [31], and that knockdown of CIC induces the transcription of ETV1, ETV4 and ETV5 in melanoma cells [29]. Moreover the transcriptional regulation by ETS transcription factors is important in human cancer development (S7 Fig). Their expression is induced in many tumors and cancer cell lines. For example, ERG, ETV1, and ETV4 can be upregulated in prostrate cancers [55], and ETV1 is upregulated in post gastrointestinal stromal tumors [56] and in more than 40% of melanomas [57]. In addition, the mRNA expression of these ETS genes was correlated with ERK activity in melanoma and colon cancer cell lines with activating mutations in BRAF (V600E), such that their expression decreased upon MEK inhibitor treatment [58]. Furthermore, overexpression of the oncogenic ETS proteins ERG or ETV1 in normal prostate cells can activate a Ras/MAPK-dependent gene expression program in the absence of ERK activation [59]. These cancer studies imply that there is an unknown factor that links Ras/Mapk activity to the expression of ETS factors, and that some of the human ETS factors might act without MAPK phosphorylation, as does *Drosophila* PntP1. Combining our knowledge of Cic with what was previously known about CIC in tumor development, we propose that CIC is the unknown factor that regulates ETS transcription factors in Ras/MAKP-activated human tumors.

In summary, our study has elucidated a mechanism wherein Cic controls the expression of the cell cycle regulators *stg (Cdc25)* and *Cyclin E*, along with the Ets transcription factor Pnt, and perhaps also Ets21C, by directly binding to regulatory sites in their promoters and introns. Using genetic tests we show that these interactions are meaningful for regulating stem cell proliferation. Therefore, we suggest that human CIC may also lead to the transcriptional induction of cell cycle genes and ETS transcription factors in RAS/MAPK activated- or loss-of-function-CIC tumors such as brain or colorectal cancers.

Materials and Methods

Drosophila stocks and transgenes

esg^{ts}: *esg-Gal4/Cyo; tubGal80^{ts} UAS-GFP/TM6B* [60]
esg^{ts} F/O: *esg-Gal4 tubGal80^{ts} UAS-GFP/Cyo;UASflp>CD2>Gal4/TM6B* [8]

Tub^{ts}: *tub-Gal4; tubGal80^{ts}/TM3*, *ser* [61](provided from Valeria Cavaliere lab)

esg^{ts}; Su(H)-Gal80: *esg-Gal4-UAS-2XEYFP; Su(H)GBE- Gal80*, *tub-Gal80^{ts}* (Gift from Steven Hou's lab)

UAS-λTOP/FM7 [37]

UAS-RAS^{v12s35} [38]

UAS-Ras RNAi [11]

UAS-Egfr RNAi [11]

UAS-cic-RNAi/Cyo (VDRC KK103805)

UAS-cic-RNAi/Cyo (VDRC KK103012)

UAS-pnt.P1 (Bloomington Drosophila Stock Center 869)

UAS-pnt.P2 (Bloomington Drosophila Stock Center 399)

UAS-pnt-RNAi (Bloomington Drosophila Stock Center 31936)

UAS-pnt-RNAi (Bloomington Drosophila Stock Center 35808)

UAS-yan-RNAi (Bloomington Drosophila Stock Center 26759)

UAS-yan-RNAi (Bloomington Drosophila Stock Center 34909)

UAS-yan-RNAi (Bloomington Drosophila Stock Center 35404)

UAS-Ets21C-RNAi (VDRC KK103211)

FRT82B cic^{fetu6} / TM3, *Sb*, *Se* (gift from Jimenez lab, Barcelona)

w; cic^{fetT6} /TM3, *Ser* (gift from Nilson lab, Canada)

w; cic^{fetE11} / TM6b (gift from Nilson lab, Canada)

w; +; UAS-cic-HA

w; UAS-cic-HA; +

w; +;UAS-cic $^{\Delta C2}$-HA

w; UAS-cic $^{\Delta C2}$-HA; +

FRT82B pnt $^{\Delta 33}$ [45,47] (gift from Joseph Bateman lab, Wolfson Centre for Age-Related Diseases)

FRT82B pnt $^{\Delta 78}$ [45,47] (gift from Joseph Bateman lab, Wolfson Centre for Age-Related Diseases)

FRT82B pnt $^{\Delta 88}$[45,47] (gift from Joseph Bateman lab, Wolfson Centre for Age-Related Diseases)

Generation of transgenic flies

The *cic $^{\Delta C2}$* was amplified from the *pCasper4—cic $^{\Delta C2}$* plasmid. The *cic* or *cic $^{\Delta C2}$* cDNAs were inserted into pUASg-attB-HA [62] vector and used to generate transgenic flies. To generate *UAS-cicDam* transgenic flies, Cic was amplified from a cDNA library prepared from midgut. This *cic* cDNA was inserted into the pUASTattB-LT3-NDam plasmid (from Andrea brand lab), and transgenics were produced.

Ectopic expression

Ectopic expression of transgenes in the adult midgut was achieved using the temperature sensitive inducible UAS-Gal4 system [63], TARGET. Crosses were set up and maintained at 18°C, the permissive temperature. 3–7 day old flies were shifted to 29°C for different times as indicated.

Bacterial infection

Gut infections were performed by feeding flies live *P.e.* in 5% sucrose on Whatman filter paper and yeast paste at 29°C.

Clonal analysis

The MARCM system was used to generate ISC clones. In order to reduce heat shock dependent stress, the clones were induced by heat shocking 3–5 days old flies at 34°C for 20 minutes. The heat shocked flies were then kept at 25°C. Clone size was measured after 10, 20, 30 days of clone induction. The size of the clones was quantified by Fiji software measuring GFP^+ area from z-projected confocal microscopy images.

Immunohistochemistry and microscopy

Female adult flies were dissected in 1×PBS. Midguts were fixed in 1×PBS with 4% paraformaldehyde for 30 minutes at room temperature. Samples were washed in PBS with 0.1% X-100 (PBST) for 3x10 minutes each. Then the tissues were blocked in PBS with 0.1% X-100, 2.5% BSA, 10% NGS for at least 30 min at room temperature. All samples were incubated with primary antibody overnight at the following dilutions: rat anti-HA (1:200; Roche), guinea pig anti-Cic (1:1000, generated by author), rabbit anti-PH3 (1:1000, Millipore). After washing 3 times 10 minutes each in PBST, samples were incubated with secondary antibodies for at least 2 hours at room temperature at a dilution of 1:1000. DNA was visualized with DAPI (0.1mg/ml, Sigma), diluted 1:200. Images of Figs 1A–1B and 2E–2H were acquired by Delta vision microscope and the rest of the fluorescence images were taken by Leica SP5 confocal microscope. Images were then processed using Fiiji and Adobe Photoshop software.

RT-qPCR

RNA was extracted from 10–12 female midguts using the RNAeasy kit (QIAGEN). RNA isolation from sorted cells was performed as previously described [64] and 100ng RNA (non-amplifed) used for reverse transcription. cDNA was synthesized by QuantiTect reverse transcription kit (QIAGEN). RT-qPCR was performed on a Light Cycler 480 II using SYBR Green I (Roche). Experiments were performed in biological triplicate. Relative fold differences in expression level of target genes were calculated as ratios to the mean of the reference genes rp49 [65] and tubulin [23]. Primer sequences are given in Supplementary Material and Methods.

RNA-Seq and data analysis

RNA isolation and amplification from sorted cells was performed as previously described [64]. Four independent biological replicates were used for sequencing. Raw reads were checked for quality using Fastqc and subsequently aligned using Tophat2, version 2.0.9, against the Flybase genome version 6. Mapped reads were counted using HTSeq-count version 0.5.4p5 [66] with mode „union". Genes showing a cpm value below 1 in four samples per treatment were considered as poorly expressed and filtered out before conducting differential expression analysis using edgeR, version 3.2.4 [67]. Since our replicates were generated independently, we used a paired design and corrected the resulting *p*-values by the Benjamini-Hochberg method [68]. Subsequently, genes with a fold change of 1.5 and an adjusted *p*-value lower than 0.1 were considered as significantly deregulated.

DamID-Seq. *UAS-Dam* and *UAS-cicDam* transgenes were induced in esg+ cells for 24 hours at 29°C and 80 guts were dissected. Genomic DNA was extracted and methylated DNA was processed and amplified as previously described [69]. Sequencing libraries were prepared according to the protocol from Andrea Brand lab (personal communication), with the following modifications. Amplified DNA from experimental and Dam-only controls was fragmented in a Covaris-S2 then digested with Sau3AI to remove the adaptors. The Truseq DNA PCR-Free

Sample Preparation kit (illumina) was used to prepare the sequencing library. The library was sent for Hiseq-2000 single-end 50bp sequencing.

DamID-Seq analysis. Raw reads were mapped to the *Drosophlia* genome (version 6.02, www.flybase.org) using Bowtie 2 [70] with default setting. After mapping, the uniquely mapped reads were extended 300 base pair (bp) toward 3' prime, and the genome were segmented with a unit of 75 bp window. Then the number of reads falling into each window was quantified. Normalization factors were computed based on the assumption that the mean log ratio between two experiments is equal to 1. The log ratio between treatment and background was observed to follow a normal distribution; the mean and standard deviation were estimated by using the fitdistr function in R (www.r-project.org). The statistical significance of the enrichment was then computed using the estimated mean and standard deviation. Using a sliding window approach, a binding site was called, where at least 4 continuous units (4x75bp windows) had a significant enrichment (p-value<0.01). The false discovery rate of the binding sites was calculated similarly to the previous publication [71].

Annotation of the peaks was carried out using two approaches. The first is based on the summit of the peak. The distance between the summit and the closest gene transcription start site (TSS) was computed with the gene orientation in consideration, and the closest gene was then assigned to the peak. The second approach is based on the entire peak. If the gene is found to be overlapping with the peak, then the gene is associated with the peak.

"TGAATG[AG]A" was searched in the peaks. The total number of occurrences was quantified, as well as the number of peaks that contained the searched pattern. In order to estimate the significance of the pattern, the background was generated by randomly moving the peaks in the genome for 1000 times. The occurrence of the pattern in the random sequences was then fitted to a Negative Binomial distribution by using the fitdistr function in R (www.r-project.org). The p-value was computed using the pnbinom function in R (www.r-project.org). To assess the association between the RNA-Seq and DamID-Seq hits, genome wide genes were ranked by fold change or absolute expression change, and corresponding number of genes which has binding sites were calculated by a moving sum (window size = 500). The absolute change is defined as the treatment value minus the background value.

***in vitro* DNA binding assays.** EMSA experiments were conducted using derivatives of Cic^{mini}, a minimal Cic protein that is functional in the embryo [25]. Wild-type and HMG-box mutant products were synthesized with the TNT T7 Quick Coupled Transcription/Translation System (Promega); the HMG-box mutant construct lacks the peptide sequence ILGEWW. DNA probes were amplified by PCR using primers carrying Not I restriction sites, digested with Not I and end-labeled with ^{32}P-dCTP and Klenow Fragment, exo- (Thermo Scientific). Binding reactions were carried out in a total volume of 20 ml containing 60 mM Hepes pH 7.9, 20 mM Tris-HCl pH 7.9, 300 mM KCl, 5 mM EDTA, 5 mM DTT, 12% glycerol, ~1 ng probe, 1 mg poly (dI-dC), 1 mg BSA and 1 ml of programmed or non-programmed (control) TNT lysate. After incubation for 20 min on ice, complexes were resolved on 5% non-denaturing polyacrylamide gels run in 0.5X TBE at 4°C, and visualized by autoradiography.

RT-qPCR primers

Rp49 –Forward: TCGATATGCTAAGCTGTC
Rp49 –Reverse: GGCATCAGATACTGTCCCTTG
β-tubulin-Forward: ACATCCCGCCCCGTGGTC
β-tubulin-Reverse: AGAAAGCCTTGCGCCTGAACATAG
pnt-Forward: ACGCCCTATGATGCTCAATC
pnt-Reverse: TATCCAGACCCAAGGTGCTC

pntP1-Forward: CGACTGCGAACAATCTGGT
pntP1-Reverse: TTGCTGGTGTTGTAGCCTGT
pntP2-Forward: TTAGCCAATTGAACGGCATC
pntP2-Reverse: GCACAGATCCTTGCATCCAT
Ets21C-Forward: CCGGGCACTCAGGTACTACT
Ets21C-Reverse: CATACTGGAGGCCGGATCT
aos-Forward: AGAACCCATGGCTTACATGC
aos-Reverse: CGTCGCGGGTGTTAAGTTAC
yan-Forward: CTGCTGGTCATCGTGCTTAG
yan-Reverse: GACCTCAGTGTGAGCAGCAA
stg-Forward: CAGCATGGATTGCAATATCAGTA
stg-Reverse: CAACGTCGTCGTCGTAGAAC
CycE-Forward: ACAAATTTGGCCTGGGACTA
CycE-Reverse: GGCCATAAGCACTTCGTC

Supporting Information

S1 Fig. Summary of cross-cancer genetic aberrations for human CIC. The figure was reproduced from the cBioPortal for Cancer Genomics web page and modified to show only cancers with >3.3% alteration frequency. Asterisks mark colorectal cancer data from Genentech [35].
(TIF)

S2 Fig. Cic regulates ISC proliferation. (A, B) RNAi-mediated depletion of Cic in ISCs and EBs using the esg^{ts} system. *esg*+ progenitor cells (green), PH3+ (red) nuclear DNA (blue). (A) Control adult midgut (B) Cic knock down midgut after 4 days induction 29°C. Scale bars represent 50μm. (C-H) Cic mutant clones were analyzed using the MARCM system. ISC clones (green), DNA stained with DAPI (blue). Control (C, E, G) and mutant (D, F, H) ISC clones were induced with the MARCM system and examined 10 days, 20 days and 30 days later. Mutant ISCs divided faster and generated bigger clones. Scale bars represent 100μm.
(TIF)

S3 Fig. Cic function has both cell autonomous and non-cell autonomous effects on ISC proliferation. (A, B) RNAi-mediated depletion of Cic in ISCs using the Dl^{ts} system. ISCs are marked by GFP (green). Sample was also stained with anti-PH3 to detect mitoeses (red) and DAPI to detect nuclear DNA (blue). (A) Control adult midgut, (B) Cic depleted midgut after 4 days induction at 29°C. Dramatic increases in the number of GFP positive cells were observed in *cic* depleted midguts, as was large increase in ISC mitoses. (C, D) RNAi-mediated depletion of Cic in EBs using the $Su(H)^{ts}$ system. EB cells are marked by GFP (green). Samples were also stained with anti-PH3 (red) and DAPI (blue). (C) Control adult midgut (D) Cic depleted midgut after 4 days induction 29°C. Increases in the number of GFP positive cells and mitoses were observed in *cic* knockdown midguts. (E) Midguts as in A-C were scored for PH3+ cells after 4 days of induction of *cic-RNAi* in ISCs or EBs. (F) After 4 days induction of *cic-RNAi* in ISCs or EBs, midguts were scored for GFP+ or GFP- mitotic cells. Most mitotic cells were GFP+ when *cic-RNAi* was induced in ISCs using the Dl^{ts} system, whereas in midguts in which *cic* was depleted in EBs, most of the mitotic cells were GFP- and likely ISCs. This indicates a non-cell autonomous effect. (G) Midguts were scored for PH3+ cells after 4 days of induction of *cic-RNAi* using the esg^{ts} system, which targets gene expression to ISCs and EBs. Dramatic increases in the number of GFP positive cells were observed in *cic* knockdown midguts as was a large increase in ISC mitoses. Statistical significance was determined by Student's t test ($^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$, $^{****}p<0.0001$). Error bars in each graph represent standard deviation.

Scale bars represent 20μm.
(TIF)

S4 Fig. Identification of Cic direct target genes in ISCs. (A) Graph showing fold change of peaks from Cic-DamID and *P.e*. infected Cic-DamID samples. (B) Cic binding sites in the *aos* locus from Cic-DamID-Seq using midgut ISCs. The black peaks are from control animals, and the grey peaks are from *P.e*. infected animals. Plot represents the log2 ratio between the Dam-fusion signal and the Dam-only signal. Red arrows point out TGAATG(G/A)A motifs. The *aos* transcription unit is shown below the graph. Yellow boxed regions indicate the ORF. (C) mRNA expression ratio change of *aos* was analyzed by qRT-PCR and normalized to *β-Tub* and *Rp49* with non-amplified mRNA from FACS-sorted progenitor cells.
(TIF)

S5 Fig. Cic directly regulates *Ets21C* and *pnt*. (A) mRNA expression heatmap of Ets transcription factors, showing fold change inductions from RNA-Seq data from whole midguts upon 6 hours *P.e*. infection. (B) Cic binding sites in the *Ets21C* locus from Cic-DamID-Seq from esg+ cells. Black peaks are from control samples and grey peaks are from *P.e*. infected midguts. The Y-axis represents the log2 ratio of the Cic-Dam fusion signal to the Dam-only signal. Red arrows point out TGAATG(G/A)A motifs. (C) Normalized mRNA expression fold change of *pnt*, *pntP1*, *pntP2* and *Ets21C* in *cic* transheterozygous mutant midguts.
(TIF)

S6 Fig. Midgut functions of *pntP1*, *pntP2*, and *yan*. (A–C) Effect of *pntP2* overexpression on ISC proliferation. Transgene expression was induced using the esg^{ts} system at 29°C for 4 days. Samples were stained with anti-GFP (green), anti-PH3(red) and DAPI (blue) to mark DNA. (A) Control adult midgut. (B) *pntP2* overexpressing midgut. The *pntP2* over expressing midgut had more GFP+ ISCs and EBs (green). (C) *pntP2* and $cic^{\Delta C2}$ over expressing midgut. GFP positive progenitor cells were still able to proliferate in the *pnt*, $cic^{\Delta C2}$ over-expressing midgut. (D) *pnt* mutant clones analyzed by the MARCM system. The size of the clones was quantified by counting cell numbers per clone. $pnt^{\Delta 33}$ is a *pntP1* specific mutant allele, $pnt^{\Delta 78}$ is *pntP2* specific mutant allele and $pnt^{\Delta 88}$ is *pnt* null mutant allele that affect both isoforms. Only the $pnt^{\Delta 88}$ detectably suppressed clone expansion. (E) Mitotic ratio of the *pnt* mutant clones was scored by calculating the average number of mitoses in each clone. (F) Quantification of ISC mitoses (PH3 positive cells) in *pnt* and *cic* depleted midguts or *yan* depleted midguts, using esg^{ts} system. Fewer mitotic ISCs were observed in the *pnt* and *cic* double knock down midgut than in the *cic* knockdown midguts, showing that *pnt* is required downstream of *cic*. Yan depletion had no effect on ISC proliferation. (G) *pnt* mutant clones were generated in a *cic* depleted background using the MARCM system. The size of the clones was quantified by counting cell numbers per clone. Only the $pnt^{\Delta 88}$ null allele suppressed the growth of *cic*-depleted ISC cell clones. (H) *yan* expression ratio as measured by qRT-PCR in *yan*-depleted midguts, using two different *yan-RNAi* lines. Statistical significance was determined by Student's t test ($^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$, $^{****}p<0.0001$). Error bars in each graph represent standard deviation. Scale bars represent 50μm.
(TIF)

S7 Fig. Summary of cross-cancer genetic aberrations for human ETS transcription factors. The figure was reproduced from the cBioPortal for Cancer Genomics web page and modified to show only cancers with >3.3% alteration frequency. (A) Cross-cancer alteration summary for EGR (the human orthologs of *Drosophila* Ets21C). (B) Cross-cancer alteration summary for EGR (the human orthologs of *Drosophila* Pnt).
(TIF)

S1 Table. Cic binding peaks from Cic DamID-seq and *P.e.* infected Cic DamID-seq. Two lists of peaks from Cic DamID-seq and Cic DamID-Seq upon infection are included in this table. Each data sheet presents the specific genomic location of the peaks with detailed information for the Cic binding peaks such as, chromosome, Peak starting sites, Peak ending sites, Log2 fold change of CicDam/Dam-only and Summit of the peaks. (S1-1) Cic binding peaks from Cic DamID-Seq. (S1-2) Cic binding peaks from Cic DamID-Seq upon *P.e*. infection.
(XLSX)

S2 Table. Genes differentially regulated (>1.5-fold, at 90% confidence) upon *cic-RNAi*. The table included the full list of genes whose mRNA expression was significantly changed (>1.5-fold, at 90% confidence) in sorted esg^{ts} *UAS-GFP* cells expressing *cic-RNAi* for 4 days. Listed data include Flybase gene ID, gene symbol, Log2-fold-change for each gene as well as significance (*p*-value with Benjamini-Hochberg correction).
(XLSX)

S3 Table. List of genes scored as Cic direct targets in progenitor cells. List of Flybase gene IDs for genes that were both upregulated upon *cic-RNAi*, and also associated with one or more Cic binding peak from the Cic DamID-Seq analysis. In addition, the normalized Log2 fold change of the peaks following *P.e*. infection, associated with the genes was also shown in the table. Some of the peaks disappeared after *P.e*. infection, so they were marked as #N/D (not detected). Derived from the data in S1 and S2 Tables.
(XLSX)

S4 Table. Potential growth-regulatory targets of Cic, identified by RNA-Seq and DamID-Seq. Information of well-known growth promoters was listed in the table with their log2 Fold change in RNA-Seq and significance (*p*-value with Benjamini-Hochberg correction) and numbers of Cic Dam-ID peaks, found in their introns or within 5kb range of the transcription start site. Genes that have Cic binding sites are shaded grey.
(PDF)

S5 Table. Cic binding to non-protein coding RNA loci. Three Lists of the non-protein coding RNAs such as tRNA, snRNA & snoRNA and non-protein coding genes that have Cic binding sites in their loci within the 5 kb range in the transcription start site were included in this table. Each data sheet shows the specific genomic location of each peak with detailed information for the Cic binding peaks such as, chromosome, Peak starting sites, Peak ending sites, Log2 fold change of CicDam/Dam-only, FlyBase ID and Gene symbol. (S5-1) Cic binding tRNAs. (S5-2) Cic binding snRNAs & snoRNAs. (S5-3) Cic binding non-protein coding genes.
(XLSX)

Acknowledgments

We would like to thank Andrea Brand for sharing the DamID-Seq experimental and analysis protocols. We specially acknowledge Jennifer Bandura for helping with the manuscript. We are also grateful to Gerado Jiménez, Laura Nilson, Iswar Hariharan, and the VDRC and Bloomington Drosophila Stock Centers for stocks and reagents. We also thank all the Edgar lab members, in particular Norman Zielke, Jerome Korzelius and Alexander Kohlmaier for stocks and discussion. We thank David Ibberson of the Heidelberg University Deep Sequencing Core Facility, Monika Langlotz from the ZMBH FACS Facility and Man Rao from Geno-Ming bioscience.

Author Contributions

Conceived and designed the experiments: BAE YJ GJ. Performed the experiments: YJ MF JX JM. Analyzed the data: NH CG.

References

1. Herbst RS (2004) Review of epidermal growth factor receptor biology. International journal of radiation oncology, biology, physics 59: 21–26.
2. Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, et al. (2006) Epidermal growth factor receptor (EGFR) signaling in cancer. Gene 366: 2–16. PMID: 16377102
3. Krasinskas AM (2011) EGFR Signaling in Colorectal Carcinoma. Pathology research international 2011: 932932. doi: 10.4061/2011/932932 PMID: 21403829
4. Downward J (2003) Targeting RAS signalling pathways in cancer therapy. Nature reviews Cancer 3: 11–22. PMID: 12509763
5. Radtke F, Clevers H (2005) Self-renewal and cancer of the gut: two sides of a coin. Science 307: 1904–1909. PMID: 15790842
6. Amcheslavsky A, Jiang J, Ip YT (2009) Tissue damage-induced intestinal stem cell division in Drosophila. Cell stem cell 4: 49–61. doi: 10.1016/j.stem.2008.10.016 PMID: 19128792
7. Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B (2009) Drosophila intestinal response to bacterial infection: activation of host defense and stem cell proliferation. Cell host & microbe 5: 200–211.
8. Jiang H, Patel PH, Kohlmaier A, Grenley MO, McEwen DG, et al. (2009) Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the Drosophila midgut. Cell 137: 1343–1355. doi: 10.1016/j.cell.2009.05.014 PMID: 19563763
9. Jiang H, Edgar BA (2009) EGFR signaling regulates the proliferation of Drosophila adult midgut progenitors. Development 136: 483–493. doi: 10.1242/dev.026955 PMID: 19141677
10. Buchon N, Broderick NA, Kuraishi T, Lemaitre B (2010) Drosophila EGFR pathway coordinates stem cell proliferation and gut remodeling following infection. BMC biology 8: 152. doi: 10.1186/1741-7007-8-152 PMID: 21176204
11. Jiang H, Grenley MO, Bravo MJ, Blumhagen RZ, Edgar BA (2011) EGFR/Ras/MAPK signaling mediates adult midgut epithelial homeostasis and regeneration in Drosophila. Cell stem cell 8: 84–95. doi: 10.1016/j.stem.2010.11.026 PMID: 21167805
12. Xu N, Wang SQ, Tan D, Gao Y, Lin G, et al. (2011) EGFR, Wingless and JAK/STAT signaling cooperatively maintain Drosophila intestinal stem cells. Developmental biology 354: 31–43. doi: 10.1016/j.ydbio.2011.03.018 PMID: 21440535
13. Jiang H, Edgar BA (2012) Intestinal stem cell function in Drosophila and mice. Current opinion in genetics & development 22: 354–360.
14. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, et al. (2009) Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature 459: 262–265. doi: 10.1038/nature07935 PMID: 19329995
15. Sato T, van Es JH, Snippert HJ, Stange DE, Vries RG, et al. (2011) Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature 469: 415–418. doi: 10.1038/nature09637 PMID: 21113151
16. Wong VW, Stange DE, Page ME, Buczacki S, Wabik A, et al. (2012) Lrig1 controls intestinal stem-cell homeostasis by negative regulation of ErbB signalling. Nature cell biology 14: 401–408. doi: 10.1038/ncb2464 PMID: 22388892
17. Roberts RB, Min L, Washington MK, Olsen SJ, Settle SH, et al. (2002) Importance of epidermal growth factor receptor signaling in establishment of adenomas and maintenance of carcinomas during intestinal tumorigenesis. Proceedings of the National Academy of Sciences of the United States of America 99: 1521–1526. PMID: 11818567
18. Weinberg R (2014) The Biology of Cancer: Garland Science.
19. Harvey Lodish A B, Kaiser Chris A., Monty Krieger, Anthony Bretscher, Hidde Ploegh, Angelika Amon, Scott Matthew P. (2012) Molecular Cell Biology: W. H. Freeman
20. Tseng AS, Tapon N, Kanda H, Cigizoglu S, Edelmann L, et al. (2007) Capicua regulates cell proliferation downstream of the receptor tyrosine kinase/ras signaling pathway. Current biology: CB 17: 728–733. PMID: 17398096

21. Jimenez G, Guichet A, Ephrussi A, Casanova J (2000) Relief of gene repression by torso RTK signaling: role of capicua in Drosophila terminal and dorsoventral patterning. Genes & development 14: 224–231.

22. Goff DJ, Nilson LA, Morisato D (2001) Establishment of dorsal-ventral polarity of the Drosophila egg requires capicua action in ovarian follicle cells. Development 128: 4553–4562. PMID: 11714680

23. Krivy K, Bradley-Gill MR, Moon NS (2013) Capicua regulates proliferation and survival of RB-deficient cells in Drosophila. Biology open 2: 183–190. doi: 10.1242/bio.20123277 PMID: 23429853

24. Jimenez G, Shvartsman SY, Paroush Z (2012) The Capicua repressor—a general sensor of RTK signaling in development and disease. Journal of cell science 125: 1383–1391. doi: 10.1242/jcs.092965 PMID: 22526417

25. Astigarraga S, Grossman R, Diaz-Delfin J, Caelles C, Paroush Z, et al. (2007) A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. The EMBO journal 26: 668–677. PMID: 17255944

26. Grimm O, Sanchez Zini V, Kim Y, Casanova J, Shvartsman SY, et al. (2012) Torso RTK controls Capicua degradation by changing its subcellular localization. Development 139: 3962–3968. doi: 10.1242/dev.084327 PMID: 23048183

27. Roch F, Jimenez G, Casanova J (2002) EGFR signalling inhibits Capicua-dependent repression during specification of Drosophila wing veins. Development 129: 993–1002. PMID: 11861482

28. Lim B, Samper N, Lu H, Rushlow C, Jimenez G, et al. (2013) Kinetics of gene derepression by ERK signaling. Proceedings of the National Academy of Sciences of the United States of America 110: 10330–10335. doi: 10.1073/pnas.1303635110 PMID: 23733957

29. Dissanayake K, Toth R, Blakey J, Olsson O, Campbell DG, et al. (2011) ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription factors via the transcriptional repressor capicua. The Biochemical journal 433: 515–525. doi: 10.1042/BJ20101562 PMID: 21087211

30. Bettegowda C, Agrawal N, Jiao Y, Sausen M, Wood LD, et al. (2011) Mutations in CIC and FUBP1 contribute to human oligodendroglioma. Science 333: 1453–1455. doi: 10.1126/science.1210557 PMID: 21817013

31. Kawamura-Saito M, Yamazaki Y, Kaneko K, Kawaguchi N, Kanda H, et al. (2006) Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Human molecular genetics 15: 2125–2137. PMID: 16717057

32. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, et al. (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery 2: 401–404. doi: 10.1158/2159-8290.CD-12-0095 PMID: 22588877

33. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling 6: pl1. doi: 10.1126/scisignal.2004088 PMID: 23550210

34. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, et al. (2006) The consensus coding sequences of human breast and colorectal cancers. Science 314: 268–274. PMID: 16959974

35. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, et al. (2012) Recurrent R-spondin fusions in colon cancer. Nature 488: 660–664. doi: 10.1038/nature11282 PMID: 22895193

36. Lee T, Luo L (2001) Mosaic analysis with a repressible cell marker (MARCM) for Drosophila neural development. Trends in neurosciences 24: 251–254. PMID: 11311363

37. Queenan AM, Ghabrial A, Schupbach T (1997) Ectopic activation of torpedo/Egfr, a Drosophila receptor tyrosine kinase, dorsalizes both the eggshell and the embryo. Development 124: 3871–3880. PMID: 9367443

38. Karim FD, Rubin GM (1998) Ectopic expression of activated Ras1 induces hyperplastic growth and increased cell death in Drosophila imaginal tissues. Development 125: 1–9. PMID: 9389658

39. Ajuria L, Nieva C, Winkler C, Kuo D, Samper N, et al. (2011) Capicua DNA-binding sites are general response elements for RTK signaling in Drosophila. Development 138: 915–924. doi: 10.1242/dev.057729 PMID: 21270056

40. Southall TD, Gold KS, Egger B, Davidson CM, Caygill EE, et al. (2013) Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. Developmental cell 26: 101–112. doi: 10.1016/j.devcel.2013.05.020 PMID: 23792147

41. Vogel MJ, Peric-Hupkes D, van Steensel B (2007) Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nature protocols 2: 1467–1478. PMID: 17545983

42. Kohlmaier A, Fassnacht C, Jin Y, Reuter H, Begum J, et al. (2014) Src kinase function controls progenitor cell pools during regeneration and tumor onset in the Drosophila intestine. Oncogene 0.

43. Shwartz A, Yogev S, Schejter ED, Shilo BZ (2013) Sequential activation of ETS proteins provides a sustained transcriptional response to EGFR signaling. Development 140: 2746–2754. doi: 10.1242/dev.093138 PMID: 23757412

44. Klambt C (1993) The Drosophila gene pointed encodes two ETS-like proteins which are involved in the development of the midline glial cells. Development 117: 163–176. PMID: 8223245

45. O'Neill EM, Rebay I, Tjian R, Rubin GM (1994) The activities of two Ets-related transcription factors required for Drosophila eye development are modulated by the Ras/MAPK pathway. Cell 78: 137–147. PMID: 8033205

46. Brunner D, Ducker K, Oellers N, Hafen E, Scholz H, et al. (1994) The ETS domain protein pointed-P2 is a target of MAP kinase in the sevenless signal transduction pathway. Nature 370: 386–389. PMID: 8047146

47. Morimoto AM, Jordan KC, Tietze K, Britton JS, O'Neill EM, et al. (1996) Pointed, an ETS domain transcription factor, negatively regulates the EGF receptor pathway in Drosophila oogenesis. Development 122: 3745–3754. PMID: 9012496

48. Xu C, Kauffmann RC, Zhang J, Kladny S, Carthew RW (2000) Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the Drosophila eye. Cell 103: 87–97. PMID: 11051550

49. Vivekanand P, Tootle TL, Rebay I (2004) MAE, a dual regulator of the EGFR signaling pathway, is a target of the Ets transcription factors PNT and YAN. Mechanisms of development 121: 1469–1479. PMID: 15511639

50. Klaes A, Menne T, Stollewerk A, Scholz H, Klambt C (1994) The Ets transcription factors encoded by the Drosophila gene pointed direct glial cell differentiation in the embryonic CNS. Cell 78: 149–160. PMID: 8033206

51. Roux PP, Blenis J (2004) ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. Microbiology and molecular biology reviews: MMBR 68: 320–344. PMID: 15187187

52. Carriere A, Cargnello M, Julien LA, Gao H, Bonneil E, et al. (2008) Oncogenic MAPK signaling stimulates mTORC1 activity by promoting RSK-mediated raptor phosphorylation. Current biology: CB 18: 1269–1277. doi: 10.1016/j.cub.2008.07.078 PMID: 18722121

53. Carriere A, Romeo Y, Acosta-Jaquez HA, Moreau J, Bonneil E, et al. (2011) ERK1/2 phosphorylate Raptor to promote Ras-dependent activation of mTOR complex 1 (mTORC1). The Journal of biological chemistry 286: 567–577. doi: 10.1074/jbc.M110.159046 PMID: 21071439

54. Ma L, Chen Z, Erdjument-Bromage H, Tempst P, Pandolfi PP (2005) Phosphorylation and functional inactivation of TSC2 by Erk implications for tuberous sclerosis and cancer pathogenesis. Cell 121: 179–193. PMID: 15851026

55. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, et al. (2007) Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature 448: 595–599. PMID: 17671502

56. Chi P, Chen Y, Zhang L, Guo X, Wongvipat J, et al. (2010) ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours. Nature 467: 849–853. doi: 10.1038/nature09409 PMID: 20927104

57. Jane-Valbuena J, Widlund HR, Perner S, Johnson LA, Dibner AC, et al. (2010) An oncogenic role for ETV1 in melanoma. Cancer research 70: 2075–2084. doi: 10.1158/0008-5472.CAN-09-3092 PMID: 20160028

58. Pratilas CA, Taylor BS, Ye Q, Viale A, Sander C, et al. (2009) (V600E)BRAF is associated with disabled feedback inhibition of RAF-MEK signaling and elevated transcriptional output of the pathway. Proceedings of the National Academy of Sciences of the United States of America 106: 4519–4524. doi: 10.1073/pnas.0900780106 PMID: 19251651

59. Hollenhorst PC, Ferris MW, Hull MA, Chae H, Kim S, et al. (2011) Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. Genes & development 25: 2147–2157.

60. Micchelli CA, Perrimon N (2006) Evidence that stem cells reside in the adult Drosophila midgut epithelium. Nature 439: 475–479. PMID: 16340959

61. Romani P, Bernardi F, Hackney J, Dobens L, Gargiulo G, et al. (2009) Cell survival and polarity of Drosophila follicle cells require the activity of ecdysone receptor B1 isoform. Genetics 181: 165–175. doi: 10.1534/genetics.108.096008 PMID: 19015542

62. Bischof J, Bjorklund M, Furger E, Schertel C, Taipale J, et al. (2013) A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development 140: 2434–2442. doi: 10.1242/dev.088757 PMID: 23637332

63. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118: 401–415. PMID: 8223268

64. Dutta D, Xiang J, Edgar BA (2013) RNA expression profiling from FACS-isolated cells of the Drosophila intestine. Current protocols in stem cell biology 27: Unit 2F 2. doi: 10.1002/9780470151808.sc02f02s27 PMID: 24510286

65. Shaw RL, Kohlmaier A, Polesello C, Veelken C, Edgar BA, et al. (2010) The Hippo pathway regulates intestinal stem cell proliferation during Drosophila adult midgut regeneration. Development 137: 4147–4158. doi: 10.1242/dev.052506 PMID: 21068063

66. Anders S, Pyl PT, Huber W (2015) HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166–169. doi: 10.1093/bioinformatics/btu638 PMID: 25260700

67. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140. doi: 10.1093/bioinformatics/btp616 PMID: 19910308

68. Benjamini Y, and Hochberg Y (1995) Controlling the False Discovery Rate—a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 57: 289–300.

69. Papagiannouli F, Schardt L, Grajcarek J, Ha N, Lohmann I (2014) The Hox gene Abd-B controls stem cell niche function in the Drosophila testis. Developmental cell 28: 189–202. doi: 10.1016/j.devcel.2013.12.016 PMID: 24480643

70. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature methods 9: 357–359. doi: 10.1038/nmeth.1923 PMID: 22388286

71. Southall TD, Davidson CM, Miller C, Carr A, Brand AH (2014) Dedifferentiation of neurons precedes tumor formation in Lola mutants. Developmental cell 28: 685–696. doi: 10.1016/j.devcel.2014.01.030 PMID: 24631403